



OPEN

# Advanced and hybrid machine learning techniques for predicting compressive strength in palm oil fuel ash-modified concrete with SHAP analysis

Tariq Ali<sup>1</sup>, Kennedy C. Onyelowe<sup>2,3✉</sup>, Muhammad Sarmad Mahmood<sup>1</sup>, Muhammad Zeeshan Qureshi<sup>4✉</sup>, Nabil Ben Kahla<sup>5,6</sup>, Aïssa REZZOUG<sup>7</sup> & Ahmed deifalla<sup>8✉</sup>

The increasing demand for sustainable construction materials has led to the incorporation of Palm Oil Fuel Ash (POFA) into concrete to reduce cement consumption and lower CO<sub>2</sub> emissions. However, predicting the compressive strength (CS) of POFA-based concrete remains challenging due to the variability of input factors. This study addresses this issue by applying advanced machine learning models to forecast the CS of POFA-incorporated concrete. A dataset of 407 samples was collected, including six input parameters: cement content, POFA dosage, water-to-binder ratio, aggregate ratio, superplasticizer content, and curing age. The dataset was divided into 70% for training and 30% for testing. The models evaluated include Hybrid XGB-LGBM, ANN, Bagging, LSSVM, GEP, XGB and LGBM. The performance of these models was assessed using key metrics, the coefficient of determination (R<sup>2</sup>), root mean square error (RMSE), normalized root means square error (NRMSE), mean absolute error (MAE) and Willmott index (d). The Hybrid XGB-LGBM model achieved the maximum R<sup>2</sup> of 0.976 and the lowest RMSE, demonstrating superior accuracy, followed by the ANN model with an R<sup>2</sup> of 0.968. SHAP analysis further validated the models by identifying the most impactful input factors, with the water-to-binder ratio emerging as the most influential. These predictive models offer the construction industry a reliable framework for evaluating POFA concrete, reducing the need for extensive experimental testing, and promoting the development of more eco-friendly, cost-effective building materials.

**Keywords** Artificial intelligence, Machine learning, POFA concrete, Compressive strength prediction, SHAP analysis, Taylor diagram

## Abbreviations

POFA	Palm Oil Fuel Ash
CS	Compressive Strength
OPC	Ordinary Portland Cement
HSC	High Strength Concrete
UHSC	Ultra-High Strength Concrete
SCC	Self-Compacting Concrete
UHPC	Ultra-High-Performance Concrete

<sup>1</sup>Department of Civil Engineering, Swedish College of Engineering and Technology, Wah 47080, Pakistan.

<sup>2</sup>Department of Civil Engineering, School of Engineering and Applied Sciences, Kampala International University, Kampala, Uganda. <sup>3</sup>Department of Civil Engineering, Michael Okpara University of Agriculture, Umudike 440109, Abia State, Nigeria. <sup>4</sup>Department of Civil Engineering, University of Engineering and Technology, Taxila, Pakistan. <sup>5</sup>Civil Engineering Department, College of Engineering, King Khalid University, Abha, Saudi Arabia. <sup>6</sup>Center for Engineering and Technology Innovations, King Khalid University, 61421 Abha, Saudi Arabia.

<sup>7</sup>College of Engineering, Imam Mohammad Ibn Saud Islamic University (IMSIU) in Riyadh, Riyadh, Saudi Arabia.

<sup>8</sup>Vice dean, Faculty of Engineering and Technology, Future University in Egypt, New Cairo, Egypt. ✉email: kennedychibuzor@kiu.ac.ug; qureshizeeshan746@gmail.com; Ahmed.deifalla@fue.edu.eg

<b>AI</b>	Artificial Intelligence
<b>ML</b>	Machine Learning
<b>R<sup>2</sup></b>	Coefficient Of Determination
<b>R</b>	Correlation Coefficient
<b>RMSE</b>	Root Means Square Error
<b>NRMSE</b>	Normalized Root Means Square Error
<b>MAE</b>	Mean Absolute Error
<b>d</b>	Willmott Index of Agreement
<b>XGB</b>	Extreme Gradient Boosting
<b>LGBM</b>	Light Gradient Boosting Machine
<b>LSSVM</b>	Least Square Support Vector Machine
<b>ANN</b>	Artificial Neural Network
<b>BPNN</b>	Back Propagation Neural Network
<b>PSO</b>	Particle Swarm Optimization
<b>ICA</b>	Imperialist Competitive Algorithm
<b>MLR</b>	Multi-Logistic Regression
<b>FQ</b>	Full Quadratic
<b>GEP</b>	Gene Expression Programming
<b>FEM</b>	Finite Element Modeling
<b>LR</b>	Linear Regression
<b>MARS</b>	Multivariate Adaptive Regression Splines
<b>GA</b>	Genetic Algorithm
<b>GP</b>	Genetic Programming
<b>SHAP</b>	SHapley Additive exPlanations

Concrete is currently consumed in enormous quantities by the building industry, with global consumption reaching up to 10 billion tons per year<sup>1,2</sup>. Due to the increasing global population, it is anticipated that by 2050, the demand for concrete would increase significantly and may reach up to 18 billion tons yearly<sup>3,4</sup> and a sizable portion of this concrete production led to the significant generation and emission of CO<sub>2</sub>, which harms the environment<sup>5</sup>. Studies show that the cement industry alone accounts for roughly 8% of global CO<sub>2</sub> emissions<sup>6–8</sup>, highlighting the urgent need for sustainability in the construction sector<sup>9</sup>. Consequently, research into alternative materials, particularly supplemental cementitious materials (SCM), has become a major area of focus<sup>10–14</sup>. Among these, the use of waste products with high pozzolanic reactivity has been explored as a potential solution to reduce environmental impact in concrete mixtures<sup>13,15–18</sup>. One such material, palm oil fuel ash (POFA), has recently gained attention as an environmentally friendly option for producing more sustainable concrete<sup>19–23</sup>.

POFA is a byproduct of producing palm oil that is produced in reactors by burning waste materials such as fiber, palm oil shells, and empty fruit bunches<sup>24</sup>. POFA's increased pozzolanic reactivity, which is the result of heat treatment, justifies its use in concrete mixes in place of cement<sup>25,26</sup>. Because of the growing demand for POFA, palm oil factories eventually produce a significant amount of fuel ash<sup>15</sup>. While Thailand produces more than 100,000 tons of POFA year on average, the Malaysian sector produces almost 3 million tons of POFA annually<sup>27</sup>. A maximum of 2 kg of palm oil can be generated for every 8 kg of raw palm oil batch; the remaining material is classified as dry biomass<sup>14,28</sup>. When this waste residue is disposed of in an untreated manner into open space and used as landfill material, it might harm people by producing various illnesses<sup>28</sup>. When compared to standard materials, POFA is more energy-efficient and environmentally benign due to its geopolymer properties<sup>29</sup>. Thus, it is possible to achieve sustainability and cost-effectiveness by substituting POFA for a portion of cement, which may also lower transportation costs and the amount of waste dumped in landfills<sup>30,31</sup>.

The most important aspect influencing concrete's strength and durability attributes is thought to be the fineness of POFA<sup>32</sup>. Numerous scholars are investigating the relationship between concrete strength and POFA fineness and pozzolanic reactivity<sup>33,34</sup>. Several research with varying outcomes improved the characteristics of concrete<sup>35–37</sup>. According to the findings of Chindaprasirt et al.'s study<sup>38</sup>, 20% OPC replacement with POFA improved workability and mechanical performances. The research additionally documented the pattern of strength degradation resulting from the elevated water consumption requirement at 40% cement replacement by POFA. Tangchirapat et al.'s<sup>39</sup> likewise recommended an increase of 8% using a comparable type of replacement. In 10–20% of POFA replacements, the ideal strength increase pattern was assessed. When compared to cement-based concrete, high-volume POFA concrete decreased production costs by 8–12% and CO<sub>2</sub> emissions by 32–45%. It results in lower natural resource usage and energy consumption, which causes the concrete industry to refocus on more ecologically friendly and sustainable manufacturing practices<sup>40–42</sup>.

It is found that, the utilization of POFA with other SCMs or high pozzolanic<sup>43,44</sup> in high strength concretes (HSC)<sup>13,33</sup> such as geopolymer<sup>45</sup>, self-compacting<sup>46</sup>, and lightweight concrete<sup>47</sup>, it greatly increases the durability and some other properties of concrete. To produce high-strength concrete, POFA was employed as a supplemental binder by Zayed et al.<sup>26</sup>. The treatment of POFA resulted in the creation of treated POFA (T-POFA) or ultrafine POFA (UPOFA), which had positive impacts on the concrete's strength, workability, and fluid transport characteristics. According to the study, UPOFA-infused high-strength concrete had better fluid transport characteristics and greater compressive strength<sup>26,33</sup>. In separate research, Zayed et al.<sup>48</sup> found that when HSC containing ground POFA (GPOFA) and OPC was compared, the latter included inferior strength, shorter setting periods, less workability, and worse physical features. This study demonstrated that the POFA's carbon content and particle size had almost no impact on the CS of HSC provided OPC replacement was limited to a maximum of 40%. Alsubari et al.<sup>40</sup> further suggested that treated POFA could improve the newly discovered properties of SCC in place of OPC. However, when compared to the control mix that contained 100% OPC, SCC

mixtures containing 50 to 70% treated POFA initially showed reduced compressive strength. Another study by Alsubari et al.<sup>49</sup> looked at the use of modified treated POFA (MT-POFA) in SCC at several percentages (0%, 30%, 50%, and 70%) as a cement substitute. The result analysis showed that the mechanical qualities of MT-POFA concrete declined over time, but that these properties significantly improved as the cure period has risen. Moreover, in studying the effect of fineness degree of POFA, Philip et al.<sup>50</sup> used 45 and 150  $\mu\text{m}$  particle sizes of POFA with 20% OPC replacement and found that found the optimal value of 45  $\mu\text{m}$  particle size which increased the CS. Notably, the results imply that POFA fineness optimization and its absorption into concrete greatly improve strength and durability, which makes it a perfect fit for environmentally friendly building methods.

Advances in artificial intelligence (AI) have led to the application of machine learning (ML) techniques for the prediction of diverse mechanical and physical characteristics of concrete<sup>51</sup>. Many machine learning techniques, like as clustering, regression, and classification, can be applied to quantify a variety of additional features to produce dependable compressive strength estimates<sup>52–54</sup>. Gene expression programming (GEP) was employed by Javed et al.<sup>55</sup> to forecast the CS of concrete composed of sugarcane bagasse ash. Moreover, ANN was used by Getahun et al.<sup>56</sup> to forecast the lifespan of concrete that contains leftover components. Moreover, ANNs, ANNs with combined inputs and ANNs with nature inspired optimizers such as particle swarm optimization (PSO), genetic algorithms (GA), beetle algorithms and ant colony optimization have been proposed in the literature in recent years. Combining several soft computing techniques can result in the development of better prediction models, such as artificial bee colony algorithms, PSO, imperialist competitive algorithm, and ANN mixed with GA. These combinations have demonstrated encouraging outcomes in resolving challenging issues, suggesting that these methods could be applied to tasks involving prediction and optimization<sup>57–60</sup>. In this regard, Yasmina et al.<sup>61</sup> used hybrid models that included ANN with additional optimization approaches including PSO and GA to forecast the CS of POFA concrete. Their findings demonstrated that the hybrid ANN models performed better than standalone ANN and other optimization strategies, delivering increased predicted accuracy. Furthermore, a few well-known machine learning techniques, like bagging, support vector machines, Extreme Gradient boosting (XGB), and hybrid machine learning models, are frequently employed extensively in huge amounts of data of CS prediction of SCMs materials<sup>62</sup>. These tree-based machine learning models are popular ensemble techniques<sup>63,64</sup>. The mechanical properties of many recently developed advanced concrete types and HSC<sup>65,66</sup>, such as self-healing concrete<sup>67</sup>, fiber-reinforced rubberized recycled aggregate concrete<sup>68,69</sup>, UHPC<sup>70</sup>, and RCA<sup>71,72</sup>, have been predicted by several studies. Thus, creating a trustworthy database containing pertinent training and test cases is essential. This first step ensures that pertinent data is available for model training and performance evaluation. Data preprocessing, regression analysis and correlation analysis are statistical investigations carried out by gathering datasets from previous studies and extracting important information that directs the creation of sophisticated ML models<sup>73</sup>.

Therefore, to perform optimization and predictive modeling, this study offers a comparative examination of advanced machine learning and deep learning models such as Bagging, improved version of SVM i.e., LSSVM, XGB, LGBM, ANN with advance structure of BPNN optimized with Adam, GEP and hybrid XGB-LGBM model. It examines and evaluates prior studies to determine the main variables affecting POFA concrete's compressive strength, such as the amount of POFA and cement, aggregate ratios, water-to-binder ratios, dose of superplasticizer, and curing time. The study investigates the link between these characteristics and compressive strength using a heatmap to graphically represent the associations, using a vast dataset of 407 samples with six input features. The results yield five performance indices: coefficient of determination ( $R^2$ ), root mean squared error (RMSE), normalized root mean squared error (NRMSE), Willmott index (d), and mean absolute error (MAE). In addition, the model's interpretability and performance are evaluated using SHAP analysis and Taylor diagrams to confirm the results and offer a thorough comprehension of the significant parameters.

## Existing literature

Predicting the characteristics of POFA concrete, particularly its compressive strength, has not received much attention in literature. Most of the research that is now accessible has created models using their own experimental datasets, which usually include less than 100 data points. The use of ML models in various concrete types is briefly covered in this section, with particular attention paid to POFA or comparable concrete types. Table 1 presents in detail the models utilized on concrete type by researchers with obtained performance metrics.

The CS of concrete incorporating steel slag was predicted using ANN, full quadratic (FQ) models, multi-logistic regression (MLR), and M5P-tree. It is found that ANN proved to be the most accurate predictor of CS, while the FQ model was most useful in estimating the electrical resistivity of the concrete<sup>74</sup>. In the case of metakaolin-containing concrete, the M5P-tree model performed better in CS forecasting rather than LR, NLR, and MLR<sup>75</sup>. In a different investigation, the FEM was used to forecast the ultimate strength of columns, and the results were compared to both GEP model and actual data. GEP produced results that were more in line with the experimental findings than FEM<sup>76</sup>. Also, while forecasting the CS of rubber modified concrete, models such as MEP, ANN, MARS, and NLR were applied, with ANN appearing as the most successful<sup>77</sup>. The M5P-tree model outperformed LR, MLR, FQ, and other techniques in terms of predicted accuracy for rubberized SCC<sup>78</sup>. Models including LR, FQ, M5P-tree, and ANN were created to estimate the CS of HSC; the findings showed that the ANN model produced the most accurate forecasts<sup>79</sup>. To estimate the CS of fly ash-modified concrete, ANN model was improved by adding PSO and the imperialist competitive algorithm (ICA); the ANN-PSO model yielded the most accurate findings in this regard<sup>80</sup>.

Recently, researchers have tended to utilize advanced AI techniques. In this regard, ANN, ANNX, PSO, GA, and other machine learning algorithms were used by Kellouche Y et al.<sup>61</sup>. utilized 249 sample dataset and employed ANN, ANNX, PSO, and GA to forecast the CS of POFA enriched concrete. They found that the ANN models outperformed PSO and GA achieving the best predictive accuracy. Moreover, Alahmari et al.<sup>68</sup> used bagging, GEP, and ANN models to predict the air-cooled CS of rubberized concrete while taking temperature,

Ref	Type of Concrete	Models applied	No. of Inputs	Dataset	Performance Metrics			Best Model
					R <sup>2</sup>	RMSE	MAE	
80	Fly ash Based concrete	ANN PSO ANN—ICA			0.838 0.877	-	-	ANN-PSO
74	Waste steel slag	ANN MLR M5P FQ	6	338	0.986 0.868 0.912 0.869	1.34 4.14 3.39 4.10	1.05 3.21 2.45 3.18	ANN
77	Waste Tire Rubber	NLR MEP ANN MARS	8	135	0.944 0.951 0.979 0.982	6.2 5.7 3.7 3.5		MARS
76	HSC	GEP	10	32	Error < 14%			
78	SCC	M5P FQ MLR LR	9	436	0.91 0.87 0.42 0.67	4.22 5.22 12.42 7.95		M5P
79	UHPC	LR ANN M5P FQ	10	192	0.81 0.93 0.88 0.86	8.72 4.97 6.84 8.50		ANN
68	Rubberized concrete	GEP ANN Bagging	4	324	0.982 0.984 0.968	0.918 0.867 1.211	0.730 0.621 0.928	ANN
61	POFA concrete	ANN ANNX PSO GA	6	249	0.989 0.977 0.934 0.923	0.033 0.049 0.082 0.087	0.023 0.023 0.057 0.066	ANN and ANN
81	UHPC	Hybrid XGB LGB	13	317	0.963 0.946 0.943	1.98 4.29 5.59	1.00 3.16 4.07	Hybrid

**Table 1.** Existing literature about ML models utilization in concrete properties prediction.

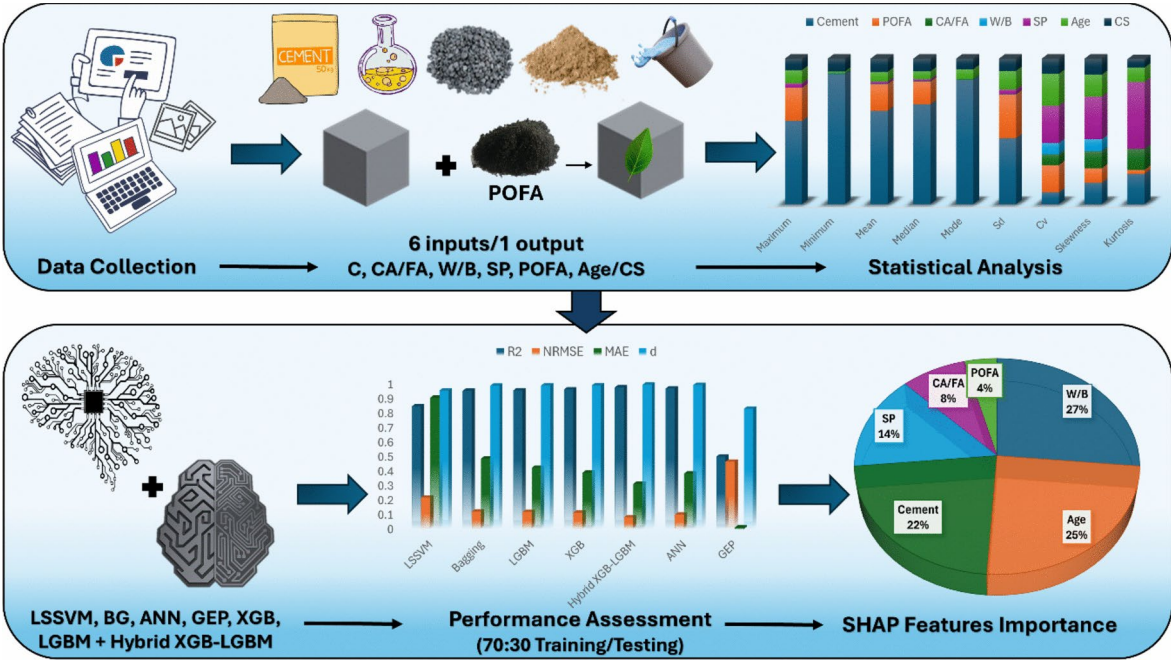
exposure time, rubber fiber content, and W/C ratio into account. ANN outperformed bagging and GEP in terms of R<sup>2</sup> value (0.984), and it also showed the lowest RMSE and MAE of all the models. Using a dataset of 626 compressive strength and 317 flexural strength data points, Das et al.<sup>81</sup> employed ML and hybrid ML models, especially XGB, LGBM, and a hybrid XGB—LGBM, to predict the CS and FS of UHPC. The hybrid XGB-LGBM model achieved the maximum accuracy in forecasting both CS and FS, with the best R<sup>2</sup> values and lowest RMSE, exceeding the separate models.

Based on a thorough review of existing literature, this study utilized advanced AI techniques, including LSSVM, Bagging, XGB, LGBM, Hybrid XGB-LGBM, BPNN, and GEP models. These models were selected due to their proven effectiveness in handling complex, non-linear relationships and their superior performance in similar predictive tasks, particularly in the context of concrete and material property prediction. LSSVM was chosen for its enhanced ability to manage non-linear data patterns in regression tasks<sup>82</sup>, while Bagging was included to improve model stability and reduce variance<sup>68</sup>. XGB and LGBM are well-known for their high accuracy and efficiency with large, structured datasets, making them ideal for predicting material properties<sup>62</sup>. The Hybrid XGB-LGBM model was selected to combine the strengths of both boosting algorithms, maximizing predictive performance. BPNN with the Adam optimizer was incorporated for its capacity to model highly non-linear relationships<sup>82</sup>, while GEP was chosen for its symbolic regression capabilities, offering interpretable results<sup>68</sup>. The selection of these models was based on their demonstrated success in similar studies and their suitability for the task of predicting compressive strength in POFA-based concrete. The details about the dataset, used parameters and models are explained in detail in the subsequent sections.

**Research methodology**  
**Description of the dataset.**

This study tries to forecast the compressive strength of POFA concrete, given the availability of experimental data on the material and its advantages to the environment and economy. Figure 1 illustrates the overall framework of the research, presenting the systematic process that includes several fundamental stages of the study pipeline. For this, a dataset of 407 samples is utilized, collected from 18 journal articles (as sourced in Table 2). Complete details, including the data sources with URLs/DOIs for accessing the original studies, are provided in the **Supplementary file (S1-2)**. Concrete’s compressive strength depends on several elements, mainly on its composition and curing conditions. Cement content, POFA dosage, water-to-binder (W/B) ratio, aggregate (CA/FA) ratio, superplasticizer, and curing time (age) are the six important variables used as model inputs. The desired output variable is the compressive strength.

To ensure the study’s reproducibility, additional details about the dataset are provided in Table 3, which presents key statistics for each variable, including the minimum, maximum, average, and standard deviation. Figure 2 illustrates the distribution and range of the variables, offering insights into the data’s spread and facilitating the identification of any unusual data points or patterns.



**Fig. 1.** Research framework outlining the key stages of the study pipeline, from dataset collection to model development and SHAP analysis.

Sr No	Ref	Author (s)	X	Sr No	Ref	Author (s)	X
1	33	Mohammed et al. (2014)	16	2	26	Zeyad et al. (2016)	36
3	84	Sata et al. (2010)	72	4	85	Tangchirapat et al. (2010)	28
5	38	Chindaprasirt et al. (2007)	8	6	17	Hamada et al. (2020)	15
7	13	Hamada et al. (2020)	12	8	14	Hasan et al. (2023)	18
9	86	Hamada et al. (2019)	12	10	87	Ul Islam et al. (2016)	24
11	88	Awal et al. (2013)	4	12	89	Jaturapitakkul et al. (2007)	13
13	90	Sata et al. (2004)	16	14	91	Zeyad et al. (2017)	21
15	92	Wan Hassan et al. (2020)	30	16	39	Tangchirapat et al. (2009)	16
17	93	Ranjbar et al. (2016)	24	18	94	Al-Mughanam et al. (2020)	42
		Total Dataset					407

**Table 2.** Collected experimental data sources. X: No. of Datapoints collected from referenced paper.

	Cement (kg/m <sup>3</sup> )	POFA (kg/m <sup>3</sup> )	CA/FA (ratio)	W/B (ratio)	SP (kg/m <sup>3</sup> )	Age (days)	CS (MPa)
Maximum	1128.00	451.20	2.99	0.95	55.00	180.00	156.00
Minimum	80.00	0.00	0.38	0.16	0.00	1.00	8.00
Mean	375.85	107.43	1.29	0.40	8.08	39.57	53.68
Median	350.00	80.00	1.28	0.35	6.24	28.00	44.90
Mode	350.00	0.00	1.27	0.35	0.00	28.00	30.10
S <sub>D</sub>	155.42	103.70	0.49	0.16	10.65	44.47	29.07
C <sub>v</sub>	0.41	0.97	0.38	0.40	1.32	1.12	0.54
Skewness	1.64	1.14	1.33	0.93	3.31	1.70	1.26
Kurtosis	5.47	0.72	3.83	-0.11	12.15	2.62	1.52

**Table 3.** Statistical characteristics of dataset.



To evaluate the robustness and dependability of the experimental results, it is imperative that the dataset spreads over the parameter values as demonstrated by Fig. 2's well-distributed frequency. Most data points for the input variables fall into one of the following ranges: [200–600] kg/m<sup>3</sup> for Cement, [0–200] kg/m<sup>3</sup> for POFA, [1.0–1.5] for CA/FA, [0.2–0.9] for W/B, [0–40] kg/m<sup>3</sup> for SP, and [0–100] days for Age. This suggests that when forecasting results for fresh data that falls between these intervals, the machine learning model will perform better. Because the distribution of data is sparse in such areas, it is anticipated that the accuracy of the model would decline for values outside of these ranges.

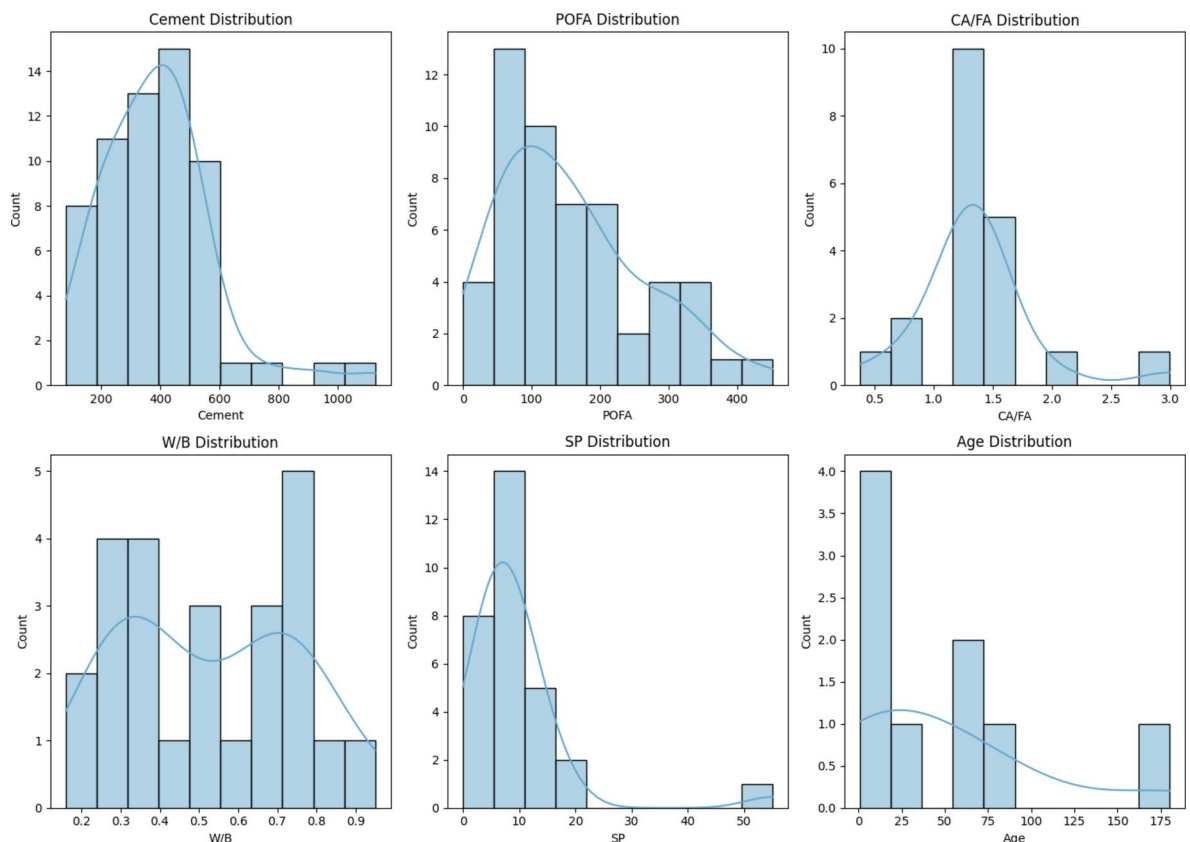
### Training and testing data

A total of 284 data points, or 70% of the dataset, are used to train the models; the remaining 123 data points or 30%, are set aside for testing.

Moreover, the occurrence of multicollinearity in machine learning models is a significant issue that arises when two or more predictor variables exhibit high levels of correlation<sup>95</sup>. Among the issues it may lead to unstable and more variance in coefficient estimates, less interpretability of the model, and a higher chance of overfitting<sup>96–99</sup>. It is intimated that the R (correlation coefficient) be maintained less than 0.8 to reduce multicollinearity<sup>100–102</sup>. Moreover, when there is a positive correlation (+1), the target variable rises together with the input variable. On the other hand, when there is a negative correlation (−1), the target variable decreases as the input variable increases<sup>96,97,103</sup>. To find strongly linked predictors, Pearson and Spearman correlation heatmap (Fig. 3) is used. This demonstrated that there was no chance of multicollinearity in the models as all the variables had  $R < 0.8$ . Notably, as shown in Fig. 3, strong positive correlations found between cement and SP with CS (0.65 and 0.7, respectively) show that these two variables significantly affect the target variable. W/B and CS have a somewhat negative association (−0.59). POFA has a negligible impact on CS; this finding may be explained by POFA's large particle size and sluggish pozzolanic activity<sup>104</sup>. Furthermore, curing time has a weakly positive connection with CS at 0.25, suggesting that although curing time affects strength, its influence is not particularly great in this instance. Although POFA and curing have a very minimal effect in this dataset, the combination of their delayed response and longer curing time implies that the compressive strength might increase over time.

### Data preprocessing

The main goal of data preprocessing is to guarantee data integrity by eliminating duplicates, blank cells and fixing mistakes. Data transformation is a crucial step in this process that entails standardization to match the scales of various variables and reduce bias in machine learning models<sup>105,106</sup>. Data normalization is a crucial preprocessing procedure that sets the input feature scale's mean to zero and standard deviation to one. By preventing any one feature from having an excessively large impact on the model due to its size, this procedure



**Fig. 2.** Statistical range of dataset.



**Fig. 3.** Multicollinearity heat map.

helps to increase the stability and performance of ML models<sup>107</sup>. “Standard Scaler” method from python library “scikit-learn” was used to normalize the dataset used in this investigation. This method scales each variable to have a unit variance and centers the data around the mean. Potential biases are reduced, and the model’s performance is improved by standardizing the data. For machine learning algorithms that are vulnerable to feature scaling, this standardization is essential since it keeps any one feature from controlling the learning process because of its size. More accurate and dependable predictions are produced when the model examines all characteristics equally, which is especially crucial for algorithms that employ gradient-based techniques or distance measures<sup>108–110</sup>.

### Advanced modeling techniques

The research uses advanced modeling techniques LSSVM, Bagging, XGB, LGBM, Hybrid XGB-LGBM, BPNN and GEP models due to their superior performance to predict the CS of POFA based concrete. The working process of these utilized models with their parameters used in study are explained below.

#### Bagging algorithm

Bagging is one of the most often used techniques for ensemble learning<sup>111</sup>. By choosing random subsets of the original dataset and combining the predictions of each individual regressor through voting or averaging, bagging regression produces a final estimate. By adding unpredictability to the model-building process, this meta-predictor technique significantly reduces variation—which often arises in black-box estimations<sup>112</sup>. It has been demonstrated through experiments on both simulated and real datasets that bagging can lead to appreciable performance gains when regression trees are employed for subset selection in linear regression and classification. The model’s estimate instability is a crucial element; bagging functions best when slight modifications to the training data result in appreciable variations to the estimator<sup>68</sup>. Although bagging can be used with any approach, it is most used in connection with decision trees and is a specific type of model averaging<sup>113</sup>. By training on a range of data subsets, bagging lessens the effect of data fluctuations, improving the performance of unstable estimators like regression trees. Combining predictions increases the model’s resiliency and accuracy<sup>112,114</sup>. The

ensemble approach is a well-established tactic because it stabilizes predictions by combining many estimators, particularly in cases when there is significant model heterogeneity<sup>115</sup>.

The equation is presented below having a dataset  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ , where  $x_i$  represents the input features and  $y_i$  is the corresponding target value. Bagging works by creating  $B$  different subsets of data  $D_1, D_2, \dots, D_B$ , each sampled with replacement from the original dataset  $D$ . For each subset  $D_b$  (where  $b = 1, 2, \dots, B$ ), a regressor  $f_b(x)$  is trained, and the final prediction  $\hat{y}$  is the average of all individual predictions from each regressor:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B f_b(x)$$

where:

- $B$  is the number of bootstrap samples (ensemble models).
- $f_b(x)$  is the prediction of the  $b$  – th regressor trained on the  $b$  – th bootstrap sample  $D_b$ .
- $\hat{y}$  is the final prediction, which is the averaged result of all  $B$  individual predictions.

The process flow of bagging is represented in Fig. 4.

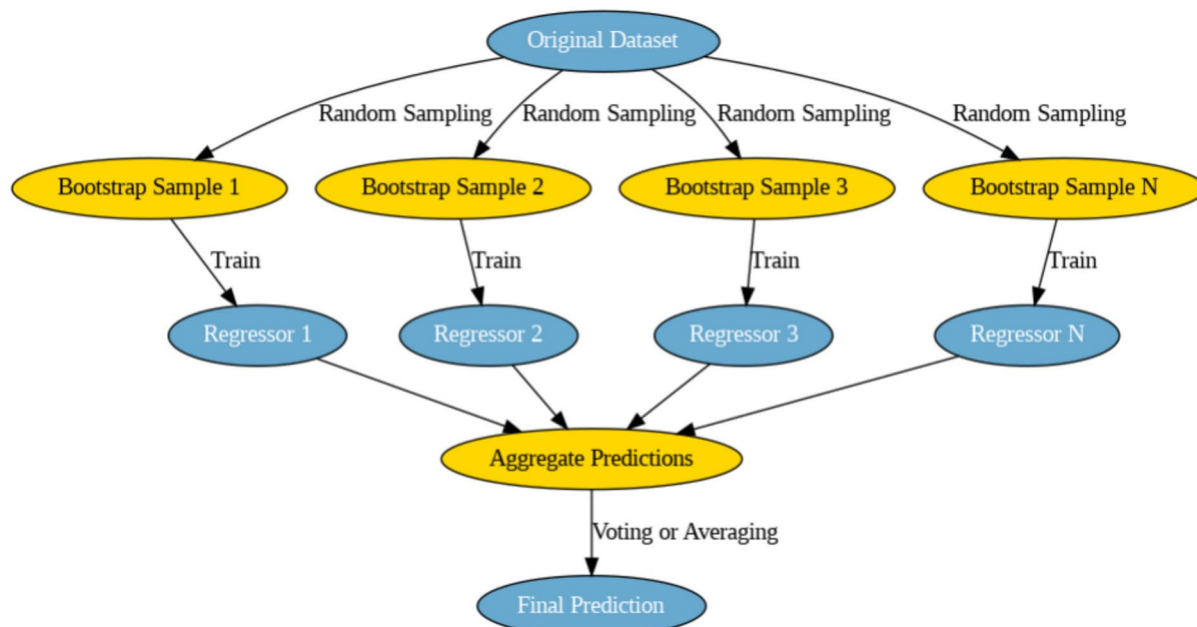
#### Least Square Support Vector Machine (LSSVM)

Suykens and Vandewalle<sup>116</sup> introduced the Least Squares Support Vector Machine (LSSVM) technique to improve Support Vector Machine (SVM) performance. LSSVM simplifies computation by addressing linear matrix problems with fewer restrictions<sup>82</sup>. One of LSSVM's main advantages over regular SVMs is its capacity to lower computing costs and minimize uncertainty when selecting structural parameters. When applied to nonlinear and small-data issues, LSSVM outperforms regular SVMs in terms of processing power. It is frequently used for jobs involving both regression and classification<sup>117,118</sup>.

But one significant disadvantage of LSSVM is that the regularization parameter ( $\gamma$ ) and the kernel function parameter ( $\sigma$ ), which is the kernel width, are critical decisions that affect the accuracy of the model. Although the model performance can be enhanced by using reconstructed input datasets with appropriate parameters, this method may induce bias when the data trends. The procedure can also be laborious and frequently necessitates prior knowledge, which can have a detrimental effect on the accuracy of the model<sup>119</sup>. The working process of LSSVM is represented on Fig. 5.

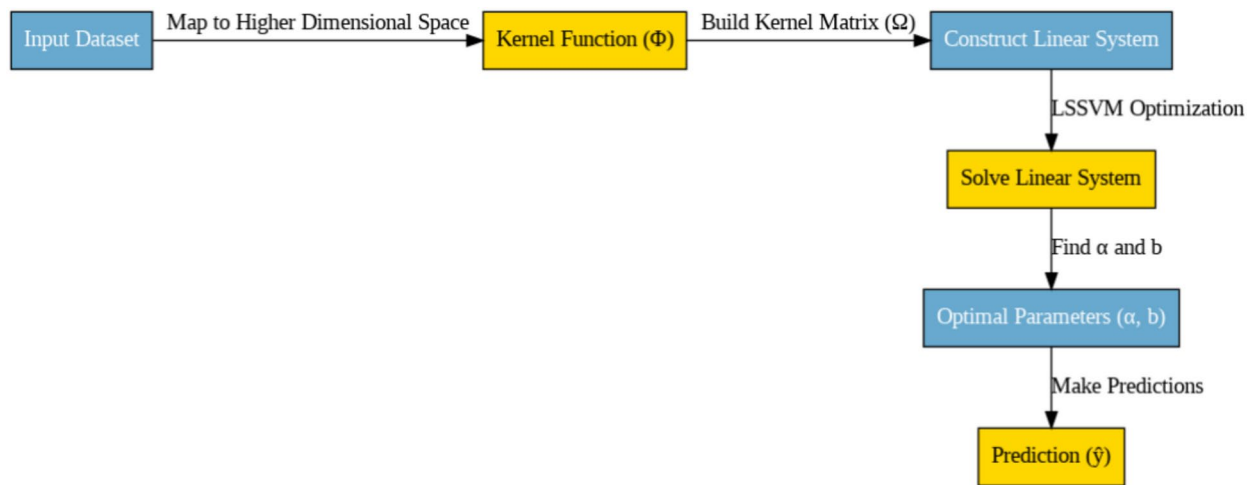
#### Light Gradient Boosting machine (LGBM)

One well-respected machine learning architecture is the Light Gradient Boosting Machine (LGBM), which processes huge datasets quickly and effectively while retaining excellent accuracy and powerful prediction powers<sup>120</sup>. As illustrated in Fig. 6, LGBM discretizes continuous data by building decision trees based on a histogram-based method. It uses leaf-wise growth techniques, building the tree leaf by leaf and dividing the



**Fig. 4.** Process flow visualization of bagging algorithm.





**Fig. 5.** Process flow visualization of LSSVM algorithm.

dataset at the leaves according to the highest data gain. By choosing the leaf with the greatest loss reduction, this technique improves the prediction accuracy of the tree<sup>121–123</sup>.

LGBM has achieved great popularity because of its effectiveness in managing enormous volumes of data while generating very accurate forecasts<sup>53</sup>.

#### *Extreme Gradient Boosting (XGB)*

Guestrin and Chen<sup>124</sup> refined and customized the LGBM technique for classification and regression tasks and called Extreme Gradient Boosting (XGB). It utilizes ensemble learning based on regression trees, which allows it to handle both regression and classification problem<sup>125</sup>. XGB significantly outperforms commonly used LGBM models in terms of overfitting and wasteful calculation since it has distributed leverages and parallel computing<sup>62</sup>.

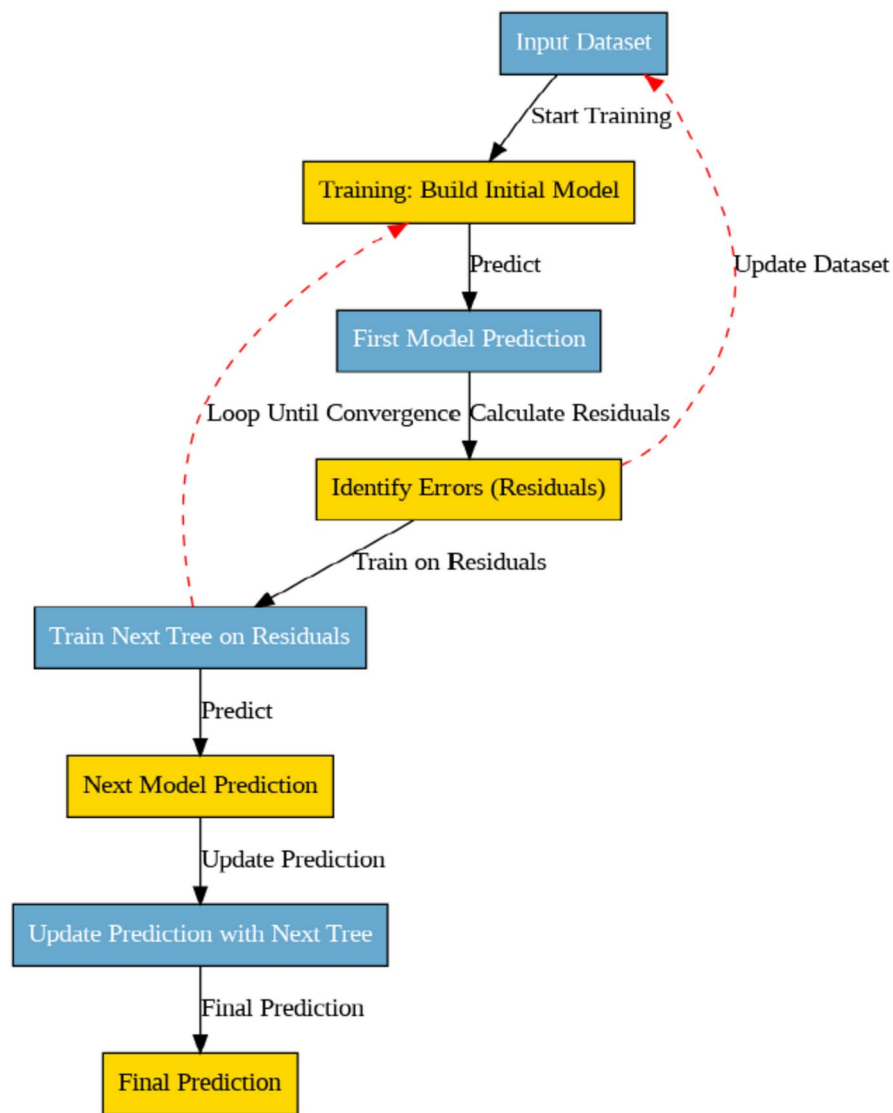
XGB is now widely used in many different sectors because of its remarkable capacity for problem-solving and low processing overhead<sup>126,127</sup>. The models in this study were forecasted using the XGB algorithm. The final prediction is produced by combining the outputs of multiple learners, and the total prediction function is calculated using the given equations. Moreover, XGB integrates techniques to reduce overfitting while maintaining high computational efficiency, allowing the algorithm to evaluate its “goodness” based on its core competencies and ultimately enhancing model performance<sup>127</sup>.

#### *Hybrid XGB-LGBM model*

The advantages of both LGBM and XGB are combined in a hybrid XGB-LGBM model to improve accuracy and performance, especially when dealing with complicated datasets<sup>81</sup>. This hybrid model was created in Google Colab using Python code and several machine learning modules. By training a first level boosting model (XGB) and using its predictions as inputs to train a second level boosting model (LGBM), for example, the two algorithms can be stacked or blended in this way. This allows one model to remedy the deficiencies of the other, enhancing overall predictive performance. The grid search technique was used to discover the optimal values for these hyperparameters, which were modified because they are critical to the accuracy and efficiency of the algorithm in this hybrid approach<sup>81</sup>. This method finds the best estimated hyper-parameters by evaluating efficacy for each possible combination of the supplied hyper-parameter. The hyperparameters used in this investigation are listed in Table 4.

#### *Artificial Neural Network (ANN)*

A computer model known as an artificial neural network (ANN) simulates the structure and functions of the human mind. This potent machine learning method is flexible and useful for problems including regression, pattern recognition, and classification. An important feature of ANN that makes them versatile in processing many kinds of data is their ability to learn complex patterns and correlations<sup>128</sup>. An ANN is made up of linked nodes arranged in layers, often known as neurons. After the input layer gets the data, it processes them through one or more hidden layers to provide the required prediction in the final output layer. Every neuron receives inputs, converts them mathematically, and produces an output that is sent to the layer next to it<sup>129</sup>. The influence of one neuron on another is determined by the connections between neurons, or weights. To minimize the discrepancy between expected and actual outputs, optimization methods are used to iteratively alter the weights during the training phase<sup>130,131</sup>. The ANN can generalize from training data and generate precise predictions on fresh, unseen data because of this learning process. Large dataset requirements, computational complexity, and the possibility of overfitting are the main issues with ANNs. When a model is overfitted to the training set, it captures noise and particular patterns that are poorly generalized to new data, which results in subpar performance on datasets that have not been seen before<sup>132</sup>. Furthermore, the quantity and quality of data have a significant impact on how effective an ANN is; inadequate or low-quality data might result in less-than-ideal model performance<sup>133</sup>.



**Fig. 6.** Process flow visualization of LGBM algorithm.

Parameters	XGB	LGBM	Hybrid XGB-LGBM
Learning Rate	0.1	0.2	0.1
Max Depth	5	4	5
Number of Estimators:	100	100	100
Number of Leaves	-	31	31
Min Data in Leaf	-	10	10
Boosting Type	-	GBDT	GBDT
L1 Regularization	-	0.1	0.1
L2 Regularization	-	0.001	0.001
Subsample	-	0.5	0.6
Random State	42	42	42

**Table 4.** Configuration of XGB, LGBM and Hybrid model.

As illustrated in Fig. 7, the input layer of the ANN model contains six variables and the output indicates CS. The subsequent values of the ANN parameters are shown in Table 5. To build a network, this study employed backpropagation, a deep learning approach.

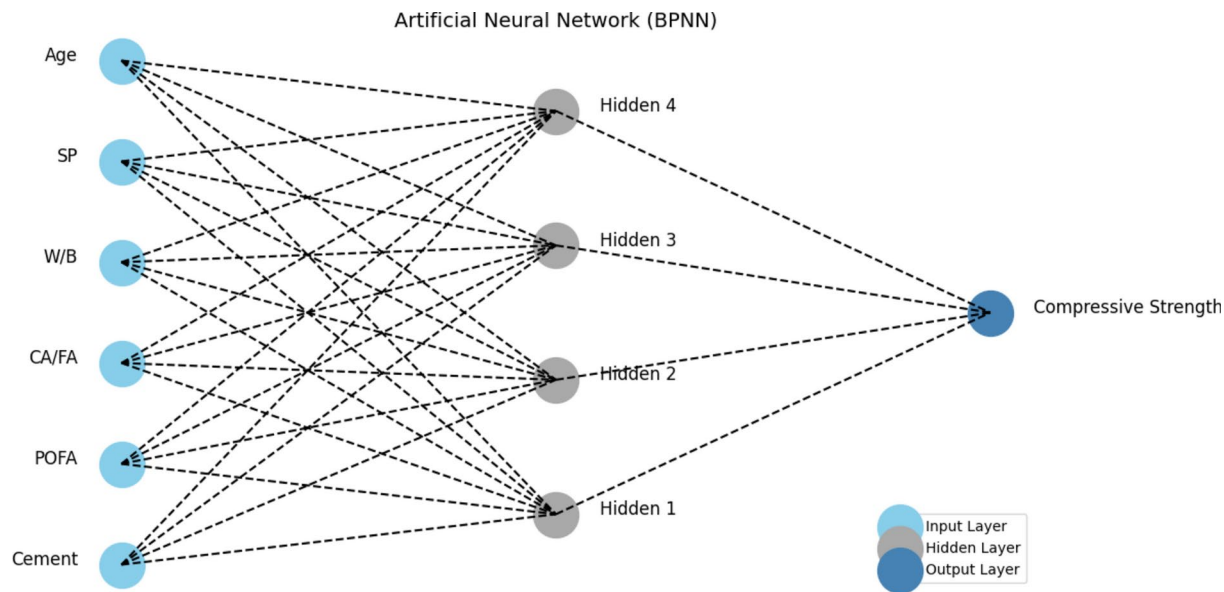


Fig. 7. Process flow visualization of ANN algorithm.

Parameter	Configuration
Type	Back Propagation Neural Network
Training function	Adam optimizer
No. of inputs	6
No. of outputs	1
No. of hidden layers	2
No. of neurons in 1 <sup>st</sup> hidden layer	64
No. of neurons in 2 <sup>nd</sup> hidden layer	64
Transfer/activation function	ReLU

Table 5. Configuration of ANN model.

This study employed backpropagation, a popular training strategy called BPNN; to improve ANN model’s learning from input data and help it produce correct predictions. It can move both forward and backward during passes. The neural network processes the input data to produce the anticipated output during the forward pass. The gradients of the model’s parameters with respect to the loss function are computed during the backward pass. To lower prediction errors, optimization methods are guided by these gradients while modifying the model’s parameters<sup>83</sup>. During training, this study employed the Adaptive Moment Estimation (ADAM) which dynamically modifies the learning rate for every parameter over time ensuring quick convergence and efficiently managing sparse gradients<sup>134</sup>. To further speed up learning, the ReLU function is employed as the transfer function in the internal hidden layers of the network since it helps prevent the vanishing gradient issue<sup>135</sup>.

Gene expression programming

The genetically enhanced population methodology was created by Ferreira<sup>134</sup>, based on The genetically enhanced population methodology was created by Ferreira<sup>136</sup>, based on GA and GP methodologies. The main applications of GA are in the optimization domain. Darwin’s theory of the survival of the fittest served as the basis for this 1960 publication by John Holland<sup>68,136</sup>. Using GP, a computer may be trained to solve issues by being informed what must be done, rather than how to execute it. This methodology creates computer programs that automatically solve problems by utilizing biological evolution theory and techniques<sup>68</sup>.

The GP approach has evolved because of three basic genetic algorithms: crossover, mutation, and reproduction. At the replication stage, a strategy for selecting which apps should discontinue use has to be put into place. During implementation, a minuscule percentage of the trees with the lowest resistance are eliminated<sup>99,137–139</sup>. GEP considerably progresses by crossing a crucial evolutionary threshold by separating the genotype and phenotype. The term “phenotype threshold” is another name for this threshold<sup>137,138</sup>.

Although GEP is a simple system, its translation mechanism is sophisticated and reliable, making it unique among artificial genotype/phenotype systems. The fact that just one gene could be passed over to the following generation was a significant constraint on GEP; since all changes occur inside a straightforward, linear framework, it was not necessary to replicate and alter the complete structure<sup>113</sup>. It is possible to evolve more

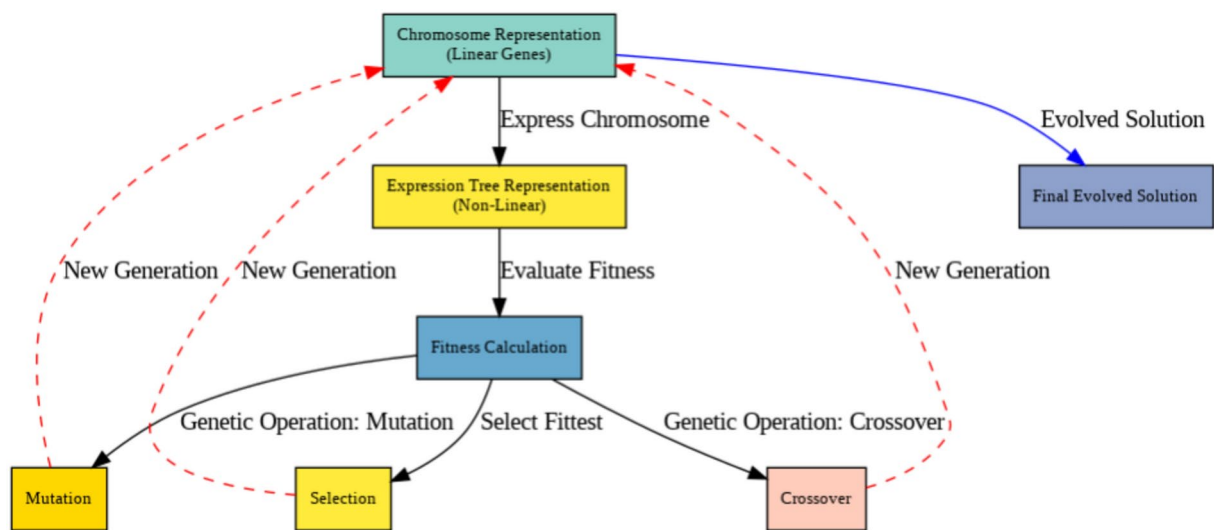


Fig. 8. Process flow visualization of GEP genetic procedures.

Sr. No	Metric Equation	Range	Unit
1	$R^2 = 1 - \frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{\sum_{i=1}^N (x_i - \bar{x})^2}$	0 – 1 (1 = perfect fit)	Unitless
2	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2}$	0 = Best Result	MPa
3	$NRMSE = \frac{RMSE}{\bar{x}} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2}}{\bar{x}}$	0 = Best Result	Unitless
4	$MAE = \frac{1}{N} \sum_{i=1}^N  x_i - \hat{x}_i $	0 = Best Result	MPa
5	$d = 1 - \frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{\sum_{i=1}^N ( x_i - \bar{x}  +  \hat{x}_i - \bar{x} )^2}$	0 – 1 (1 = perfect agreement)	Unitless

Table 6. Performance Metrics.  $x_i \hat{x}_i \bar{x}$  :  $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ ;  $N$ : Actual observed values; : Predicted values generated by the model; Mean of the observed values, : Number of observations.

complex solutions than GP because of GEP’s unique encoding method, which preserves genetic information in a linear fashion. The reason for this is that evolving programs that grow too quickly often cause GP to encounter issues. In many different fields, when traditional approaches may not be able to find the best solutions, GEP is a useful tool for predictive modeling and optimization. This can be attributed to its more straightforward depiction and evolutionary process<sup>140</sup>. Symbolic regression is one of the main advantages of GEP; it finds the optimal geometric expression to match the data without having a pre-established model structure. When the correlations between the variables are complicated or unclear, GEP’s flexibility allows it to study a larger variety of options<sup>141</sup>. Figure 8 shows a thorough flow of the genetic procedures.

Performance assessment

Every model undergoes a thorough evaluation procedure that computes many performance indicators to determine its reliability and efficacy. The Willmott index of agreement (d), coefficient of determination (R<sup>2</sup>), root mean square error (RMSE), normalized root means square error (NRMSE) and mean absolute error (MAE) are among these metrics. Consequently, as can be seen from the equations in Table 6, which give the mathematical formulas for these measurements with standard range, offer perceptive examination of the model’s predicting skills under various conditions.

Results and discussion  
Prediction accuracy

The reliability, effectiveness, and predictive capacity of ML models may all be evaluated with the use of regression analysis. A high linear connection between expected and actual values is shown by a regression coefficient (R<sup>2</sup>) that is close to 1 (especially around 0.8), which shows that the model’s predictions closely match the observed

data<sup>142,143</sup>. The  $R^2$  values of the models that were assessed in this study provide insight into the predictive power of each model.

With a  $R^2$  value of 0.976, the Hybrid XGB-LGBM model performs the best, showing the highest level of predicted accuracy and dependability. XGB model alone also shows good predictive power with an  $R^2$  value of 0.962, indicating a high degree of agreement between anticipated and actual values. With an  $R^2$  of 0.966, the BPNN (Backpropagation Neural Network) model outperforms the XGB model in accuracy. The LGBM model produces an  $R^2$  of 0.956, suggesting strong performance however it lags hybrid and BPNN. In the same manner, bagging model also produce good accuracy of 0.950.

It's interesting to note that the LSSVM model did not perform as well as the other models in this study, although still showing good predictive potential with an  $R^2$  of 0.849. Also, the GEP model performs the worst, with an  $R^2$  of 0.499, which suggests that the model's predictions and the actual data are not as well correlated. This may be due to non-linear relationships and the challenge of optimizing its evolutionary process, making it less suited for high-dimensional or intricate datasets compared to Hybrid and ANN model, which are better at capturing such complexities.

Regression graphs illustrating testing phase results are shown in Fig. 9 (a-g). By using normalized data, deviations originating from variations in the size or magnitude of variables are successfully addressed, allowing for an accurate assessment of the model's performance.

### Comparative analysis of models

Comprehensive insights into the prediction power of several predictive models are obtained by evaluating them using a variety of measures. Table 7 displays the models' comparison with respect to accuracy and error metrics.

The hybrid XGB-LGBM model is the best performer, showing greater predictive accuracy across all parameters, with the greatest  $R^2$  value of 0.976 and the lowest NRMSE (0.080) and MAE (3.11). Likewise, XGB and ANN (BPNN) exhibit outstanding performance, displaying  $R^2$  values of 0.962 and 0.968, respectively, and comparatively low MAE and NRMSE values. Based on these findings, it can be concluded that the models have good predictive power and low error rates throughout training and testing. With  $R^2$  values of 0.955 and 0.950 and low NRMSE values of 0.117 and 0.121, respectively, the LGBM and Bagging models likewise exhibit excellent performance. They nonetheless have strong predictive performance with minimal error margins, while being marginally less accurate than the Hybrid XGB-LGBM, XGB, and ANN models.

Interestingly, the GEP model, despite having a relatively low  $R^2$  value of 0.499, exhibits a very low MAE of 0.11. This paradox can be attributed to the nature of the GEP model's learning process. While the MAE suggests that GEP produces predictions with small average absolute deviations from the observed data, its low  $R^2$  and high NRMSE (0.464) indicate that these predictions do not capture the variance and overall distribution of the dataset as effectively as the other models. This suggests that while GEP may predict small deviations accurately, it struggles with more complex relationships in the data, leading to poorer overall accuracy and performance. Moreover, Fig. 10 presents a detailed comparison of the forecast and experimental CS for all models throughout with the help of error metrics. The accuracy of the Hybrid followed by ANN in predicting the CS is clearly dominant in this graphic representation, which confirms their suitability for POFA incorporated concrete CS prediction.

Moreover, the spider graph (Fig. 11) illustrates a comparative prediction power analysis of multiple models across four evaluation metrics:  $R^2$ , NRMSE, MAE, and Willmott's Index of Agreement (d) which clearly indicates that Hybrid model performs better than other models followed closely by ANN.

### Performance assessment

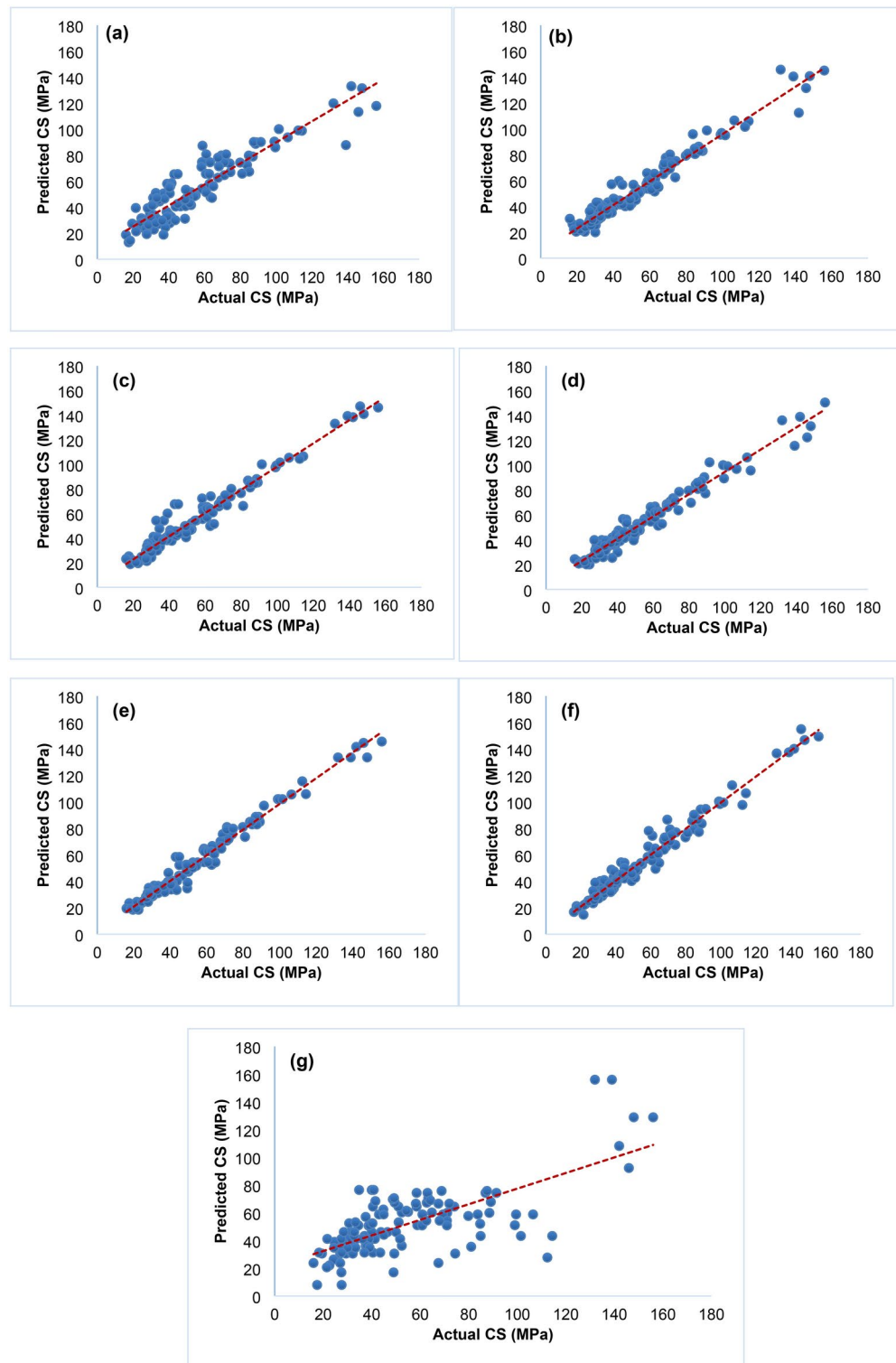
The Taylor diagram, which shows variances in standard deviation, correlation coefficient (R), and RMSE, offers a brief yet simple evaluation of each machine learning model's performance with respect to the reference dataset. These discrepancies are represented by points in Fig. 12, and the centered RMSE is indicated by contour lines, which facilitates the evaluation and ranking of the models<sup>144</sup>. The Hybrid XGB-LGBM model, which is positioned close to the ideal correlation coefficient of 1.0 and has low RMSE and standard deviation, suggesting its higher prediction accuracy, shows the closest agreement with the reference data in the diagram.

In terms of overall prediction accuracy, the ANN model falls closely behind Hybrid XGB-LGBM, but it matches standard deviation well, as indicated by its placement on the dotted line, which represents a measured standard deviation of 30. According to this location, the ANN performs well in matching the measured data's range and variability, but it is less accurate in terms of other parameters of accuracy. Conversely, the GEP model performs the worst, with a location that is farthest from the optimal line of 30, a higher predicted standard deviation, and a poorer correlation value. When compared to GEP, models such as LSSVM and Bagging exhibit intermediate performance, but still closer to reference. The models' accuracy can be ordered as follows using the Taylor diagram: Hybrid XGB-LGBM > ANN > XGB > LGBM > Bagging > LSSVM > GEP.

### SHAP Interpretation

SHAP analysis helps explain the outputs of machine learning models by assigning a value to each feature, indicating its contribution to the prediction. These values reveal how different features influence the model's outcome, providing insight into the model's decision-making process. By examining SHAP values, it becomes clear which features have the most significant impact and how they interact<sup>109</sup>. This method aids in understanding, interpreting, and validating the model through both qualitative and quantitative means (Illustrated in Fig. 13 and 14). With a mean absolute SHAP value of roughly 0.065, the water-to-binder ratio (W/B) is the most significant of the six parameters taken into consideration. This suggests that W/B has the largest influence on the anticipated CS, with larger W/B ratios (shown by red spots in the SHAP summary plot) typically translating into





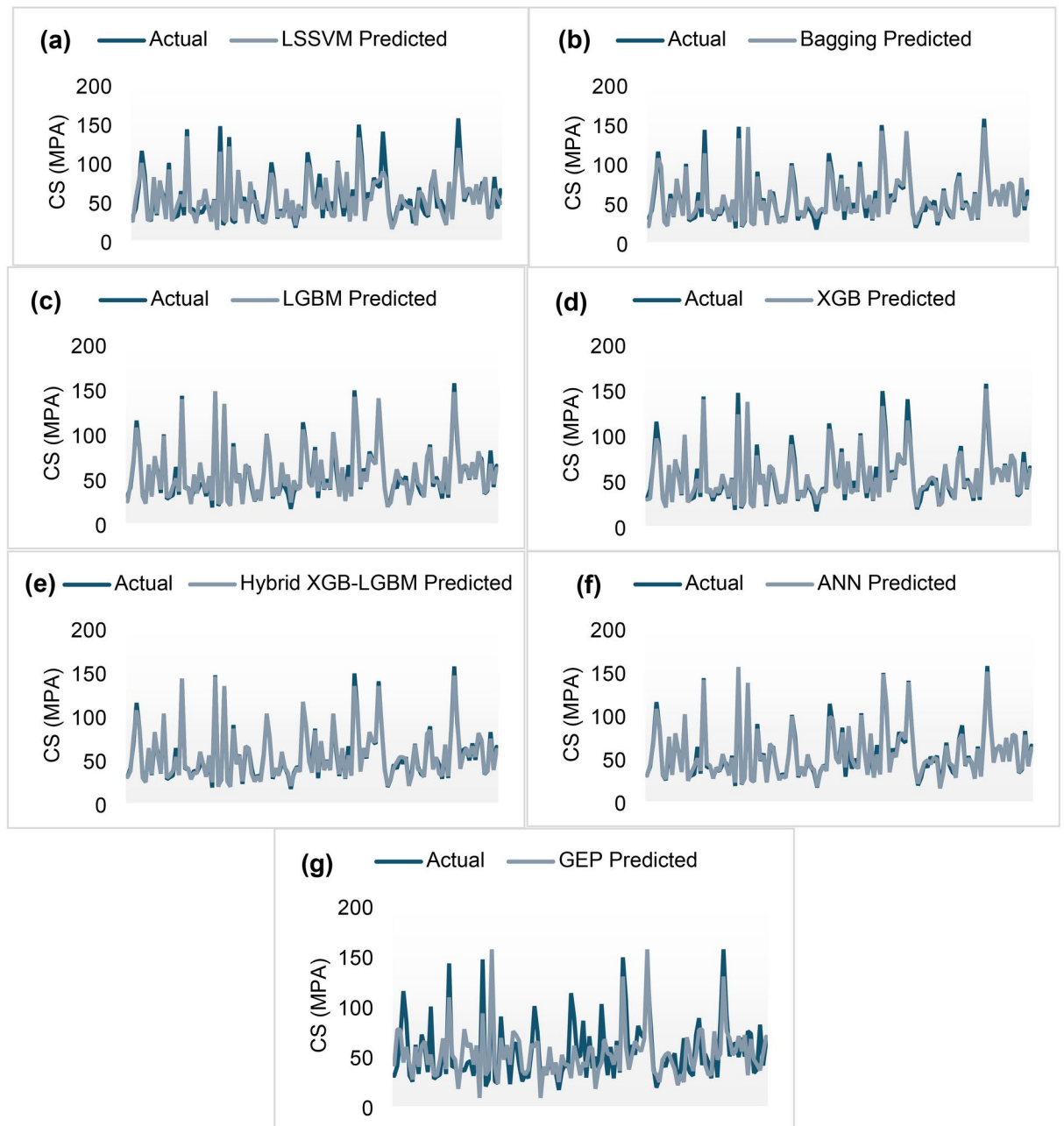
**Fig. 9.** Prediction Accuracy (a) LSSVM, (b) Bagging, (c) LGBM, (d) XGB, (e) Hybrid XGB-LGBM, (f) ANN (BPNN), (g) GEP.

estimates of higher compressive strengths. W/B has a crucial role in influencing the workability and strength of concrete, which is consistent with its substantial influence.

Age has a mean absolute SHAP value of 0.060, making it the second most influential attribute. According to the SHAP analysis, age is a crucial factor in non-linear relationships, contrary to earlier heatmap correlation studies which claimed age had little effect on the CS. Though its linear correlation is smaller, it is possible that longer curing times lead to stronger concrete. Cement ranks third, demonstrating its noteworthy contribution to

Model	R <sup>2</sup>	NRMSE	MAE (MPa)	d
LSSVM	0.845	0.217	9.05	0.954
Bagging	0.952	0.121	4.85	0.987
LGBM	0.955	0.117	4.22	0.988
XGB	0.962	0.112	3.88	0.989
Hybrid XGB-LGBM	0.976	0.080	3.11	0.994
ANN (BPNN)	0.968	0.099	3.82	0.992
GEP	0.499	0.464	11.1	0.828

**Table 7.** Models accuracy with respect to performance indicators.



**Fig. 10.** Comparison of models with error representation: (a) LSSVM, (b) Bagging, (c) LGBM, (d) XGB, (e) Hybrid XGB-LGBM, (f) ANN (BPNN), (g) GEP.

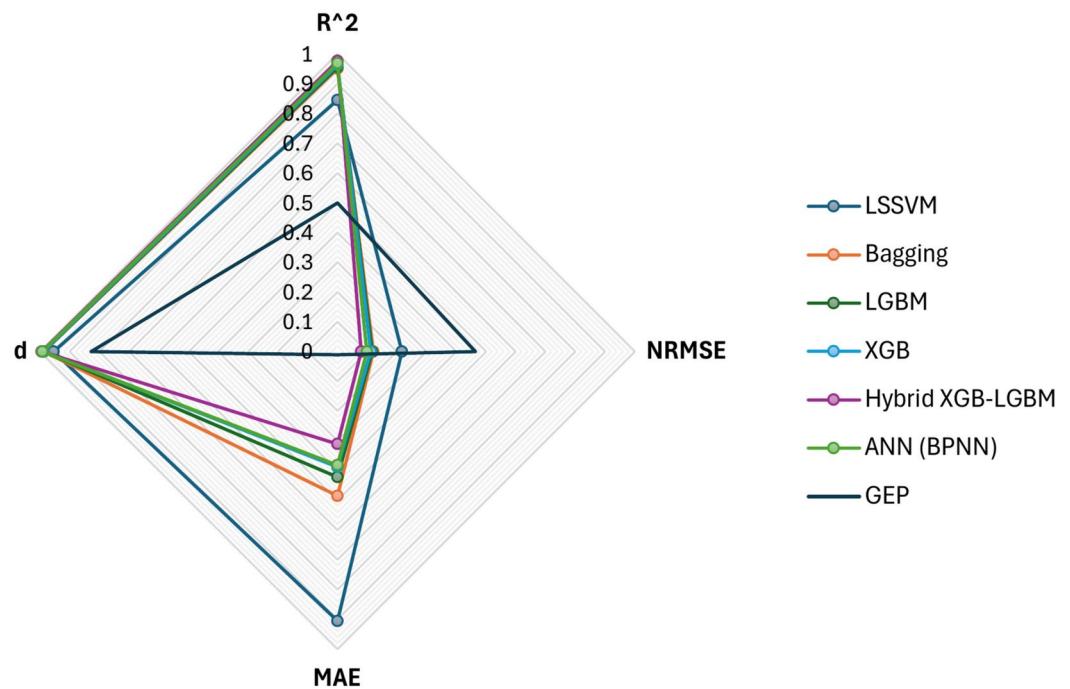


Fig. 11. Comparison of models with respect to performance indicators (Spider Graph).

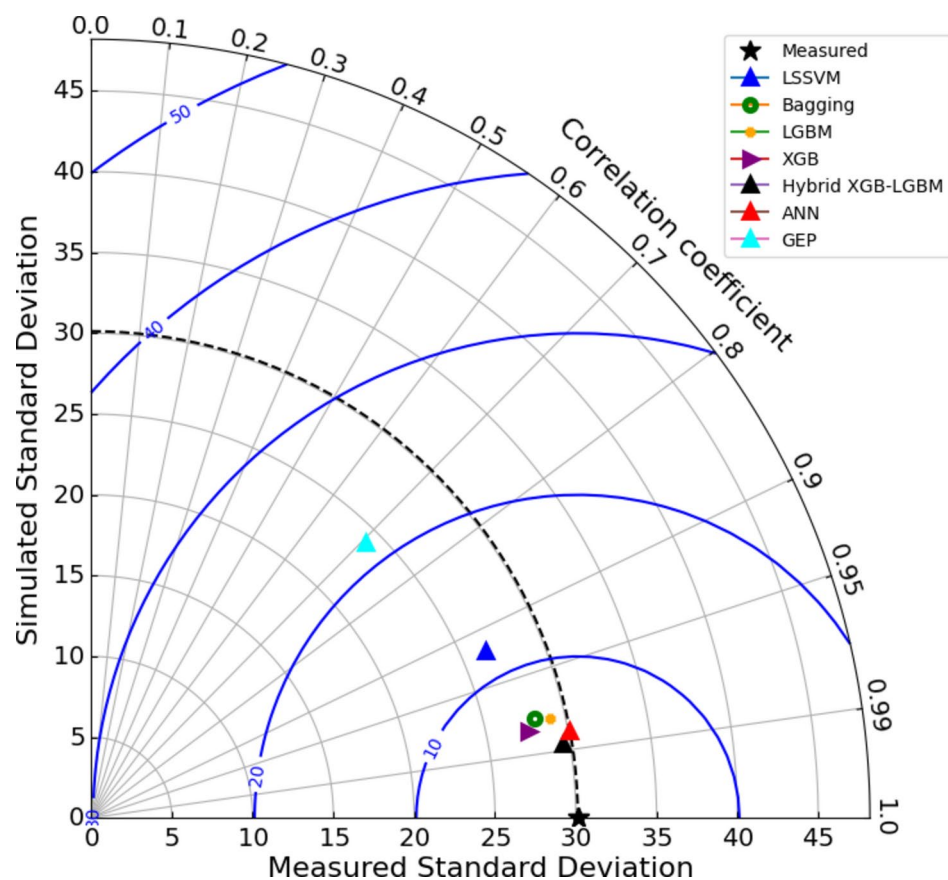
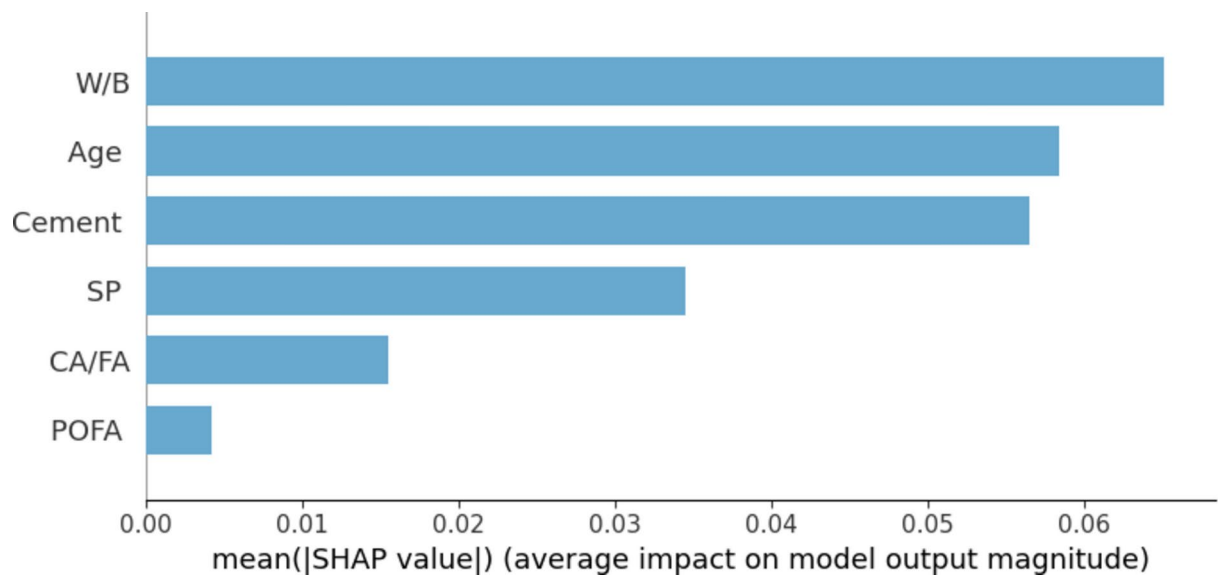
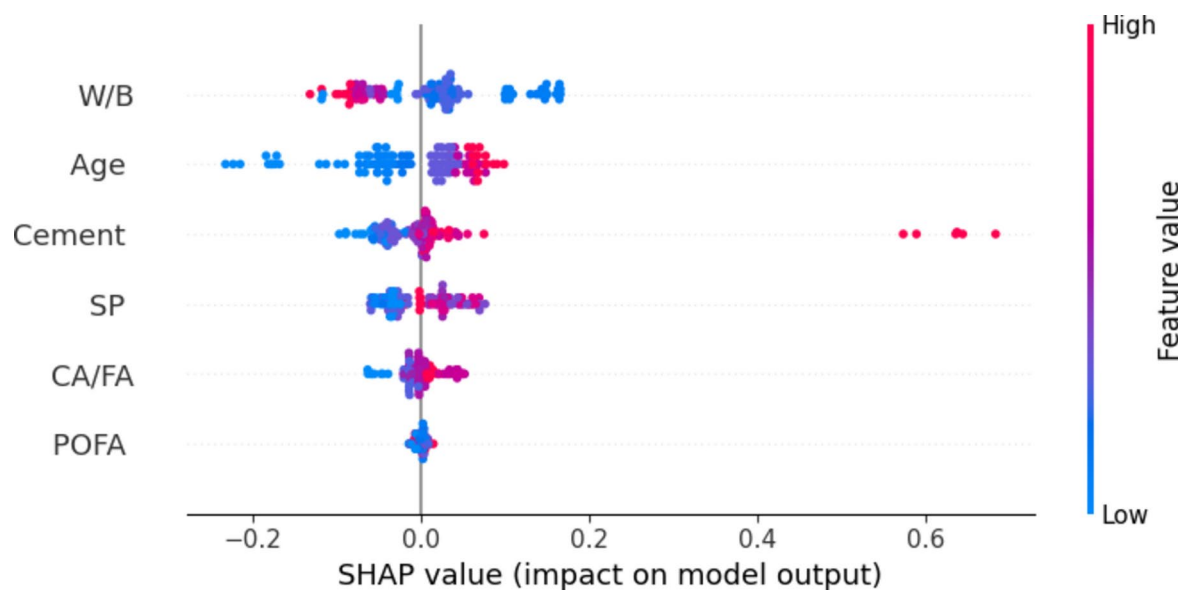


Fig. 12. Performance assessment of models in terms R, RMSE and Standard deviation.



**Fig. 13.** Feature importance using SHAP analysis.



**Fig. 14.** Summary plot of SHAP analysis.

CS with a mean absolute SHAP value of 0.055. As indicated by the positive SHAP values, a higher cement content increases the expected compressive strength which is an essential function in the process of concrete hydration. Furthermore, CA/FA and SP, with typical SHAP values of roughly 0.035 and 0.025, respectively, are ranked fourth and fifth in significance. Higher values (shown by red dots) increase compressive strength, indicating that SP has a role in both reducing water content and enhancing workability. Lower values, on the other hand, exhibit varying effects. The effects of CA/FA are inconsistent; in some situations, larger ratios increase strength, but in other situations, they decrease it. The balance between aggregate size and total mix design, which is dependent on cement content and W/B, may be the cause of this heterogeneity.

Finally, with a mean absolute SHAP value of about less than 0.01, POFA has the least influence on the model's predictions. The SHAP summary plot indicates that POFA has a largely neutral effect, with the majority of SHAP values falling close to zero which was also confirmed by correlation heat map. This suggests that the compressive strength predictions are not significantly affected by changes in POFA content. This finding may be explained by POFA's comparatively large particle size and slow pozzolanic activity, which may limit their ability to contribute to the early development of strength in the concrete mix.

## Conclusion

This study effectively addresses the challenges of incorporating POFA into concrete by utilizing advanced machine learning models such as Hybrid XGB-LGBM, ANN with structure of BPNN optimized with Adam, LSSVM, XGB, LGBM and Bagging to predict its compressive strength. These models were developed and assessed using an extensive dataset of 407 samples, demonstrating the efficacy of machine learning approaches in capturing complex correlations within the input parameters. The study divided the dataset into training and testing in the ratio of 70:30 with six inputs and validated the prediction performance of models based on  $R^2$ , RMSE, NRMSE, MAE and d. Furthermore, the model output was validated and interpreted with the assistance of SHAP analysis, which offered more insight into the feature contributions. The study's principal conclusions are summarized as follows:

- Among the evaluated models, the Hybrid XGB-LGBM emerged as the most effective, achieving the highest  $R^2$  value of 0.976, followed closely by ANN with an  $R^2$  of 0.968. XGB demonstrated good performance but marginally lesser accuracy, with an  $R^2$  of 0.962.
- While the Hybrid XGB-LGBM and ANN models showed strong predictive accuracy, GEP ( $R^2=0.495$ ) and LSSVM ( $R^2=0.865$ ) performed poorly. Specifically, GEP struggled to capture complex relationships, which made the model less effective in forecasting the compressive strength of POFA-based concrete.
- The Taylor diagram further validates the superior performance of the Hybrid XGB-LGBM model, as it is closest to the ideal correlation coefficient of 1.0 with low RMSE and standard deviation. While the ANN model closely follows, it matches the measured standard deviation well but is slightly less accurate overall. In contrast, the GEP model performs the worst, showing the highest deviation from the optimal line.
- SHAP analysis highlighted the water-to-binder (W/B) ratio as the most influential factor in determining compressive strength, followed by curing age and cement content. In contrast, POFA content showed minimal impact on the CS, aligning with its slow pozzolanic activity and larger particle size.

By using advanced ML algorithms to precisely estimate the CS of POFA-based concrete, this research greatly improves the capacity of the construction sector to use POFA in concrete. These prediction models contribute to the creation of more affordable and environmentally friendly building materials by providing a dependable framework for evaluating POFA concrete in practical applications thus reducing the need for extensive and costly experimental testing. Future research should focus on expanding the dataset and incorporating additional predictors to improve the accuracy of the models such as type of concrete and POFA particle size. Exploring advanced techniques such as hybrid models of ANN and LSSVM with nature inspired optimizers is recommended to enhance predictive performance. Additionally, applying these models to investigate other properties of concrete, beyond compressive strength, will provide a more comprehensive understanding of POFA-incorporated concrete in various construction applications.

## Data availability

All the data with referenced links is available in the supplementary file.

Received: 23 October 2024; Accepted: 4 February 2025

Published online: 10 February 2025

## References

1. Meyer, C. The greening of the concrete industry. *Cem. Concr. Compos.* **31**, 601–605. <https://doi.org/10.1016/J.CEMCONCOMP.2008.12.010> (2009).
2. Hamada, H. et al. Influence of palm oil fuel ash on the high strength and ultra-high performance concrete: A comprehensive review. *Eng. Sci. & Technol. Int. J.* **45**, 101492. <https://doi.org/10.1016/J.JESTCH.2023.101492> (2023).
3. Monteiro: Concrete: microstructure, properties, and - Google Scholar, (n.d.). [https://scholar.google.com/scholar\\_lookup?title=Concrete%3A%20microstructure%2C%20properties%2C%20and%20materials&publication\\_year=2006&author=P.%20Monteiro](https://scholar.google.com/scholar_lookup?title=Concrete%3A%20microstructure%2C%20properties%2C%20and%20materials&publication_year=2006&author=P.%20Monteiro) (accessed October 8 2024)
4. Favier, A., De Wolf, C., Scrivener, K. & Habert, G. A sustainable future for the European Cement and Concrete Industry: Technology assessment for full decarbonisation of the industry by 2050. ETH Zurich. <https://doi.org/10.3929/ethz-b-000301843> (2018).
5. Colangelo, F. et al. Innovative materials in Italy for eco-friendly and sustainable buildings. *Materials* **14**(8), 2048. <https://doi.org/10.3390/MA14082048> (2021).
6. Benhelal, E., Zahedi, G., Shamsaei, E. & Bahadori, A. Global strategies and potentials to curb CO2 emissions in cement industry. *J. Clean. Prod.* **51**, 142–161. <https://doi.org/10.1016/J.JCLEPRO.2012.10.049> (2013).
7. N. Ismael, M.G.-T.J. of E. Sciences, undefined (2016) The Effects of Lime Addition and Fineness of Grinded Clinker on Properties of Portland Cement Iasj.NetNS Ismael MN GhanimTikrit Journal of Engineering Sciences iasj.Net (n.d.) 1813–162. <https://www.iasj.net/iasj/download/8d5befb120d767bd> accessed October 8 2024
8. York: Concrete needs to lose its colossal carbon footprint - Google Scholar, (n.d.). [https://scholar.google.com/scholar\\_lookup?title=Concrete%20needs%20to%20lose%20its%20colossal%20carbon%20footprint&publication\\_year=2021&author=I.%20York&author=I.%20Europe](https://scholar.google.com/scholar_lookup?title=Concrete%20needs%20to%20lose%20its%20colossal%20carbon%20footprint&publication_year=2021&author=I.%20York&author=I.%20Europe) (accessed October 8, 2024)
9. Amin, M., Zeyad, A. M., Tayeh, B. A. & Agwa, I. S. Effect of glass powder on high-strength self-compacting concrete durability. *Key Eng. Mater.* **945**, 117–127. <https://doi.org/10.4028/P-W4TCJX> (2023).
10. Hakeem, I. Y. et al. Properties and durability of self-compacting concrete incorporated with nanosilica, fly ash, and limestone powder. *Struct. Concr.* **24**, 6738–6760. <https://doi.org/10.1002/SUCO.202300121> (2023).
11. Ghanim, A. A. J., Amin, M., Zeyad, A. M., Tayeh, B. A. & Agwa, I. S. Effect of modified nano-titanium and fly ash on ultra-high-performance concrete properties. *Struct. Concr.* **24**, 6815–6832. <https://doi.org/10.1002/SUCO.202300053> (2023).
12. Khan, M. M. H. et al. Effect of various powder content on the properties of sustainable self-compacting concrete. *Case Stud. Constr. Mater.* **19**, e02274. <https://doi.org/10.1016/J.CSCM.2023.E02274> (2023).
13. Hamada, H., Tayeh, B., Yahaya, F., Muthusamy, K. & Al-Attar, A. Effects of nano-palm oil fuel ash and nano-eggshell powder on concrete. *Constr. Build. Mater.* **261**, 119790. <https://doi.org/10.1016/J.CONBUILDMAT.2020.119790> (2020).



14. Hasan, N. M. S. et al. Eco-friendly concrete incorporating palm oil fuel ash: Fresh and mechanical properties with machine learning prediction, and sustainability assessment. *Heliyon* **9**, e22296. <https://doi.org/10.1016/j.heliyon.2023.E22296> (2023).
15. Hamada, H. M., Jokhio, G. A., Yahaya, F. M., Humada, A. M. & Gul, Y. The present state of the use of palm oil fuel ash (POFA) in concrete. *Constr. Build. Mater.* **175**, 26–40. <https://doi.org/10.1016/j.conbuildmat.2018.03.227> (2018).
16. Abdul-Rahman, M., Al-Attar, A. A., Hamada, H. M. & Tayeh, B. Microstructure and structural analysis of polypropylene fibre reinforced reactive powder concrete beams exposed to elevated temperature. *J. Build. Eng.* **29**, 101167. <https://doi.org/10.1016/j.jobe.2019.101167> (2020).
17. Hamada, H. M. et al. Effect of high-volume ultrafine palm oil fuel ash on the engineering and transport properties of concrete. *Case Stud. Constr. Mater.* **12**, e00318. <https://doi.org/10.1016/j.cscm.2019.E00318> (2020).
18. Al-Attar, A. A., Abdulrahman, M. B., Hamada, H. M. & Tayeh, B. A. Investigating the behaviour of hybrid fibre-reinforced reactive powder concrete beams after exposure to elevated temperatures. *J. Mater. Res. & Technol.* **9**, 1966–1977. <https://doi.org/10.1016/j.jmrt.2019.12.029> (2020).
19. Jokhio, G. A., Hamada, H. M., Humada, A. M., Gul, Y. & Abu-Tair, A. Environmental benefits of incorporating palm oil fuel ash in cement concrete and cement mortar. *E3S Web Conf.* **158**, 03005. <https://doi.org/10.1051/E3SCONF/202015803005> (2020).
20. Hamada, H. M., Yahaya, F., Muthusamy, K. & Humada, A. Comparison study between POFA and POCF in terms of chemical composition and physical properties-Review paper. *IOP Conf. Ser. Earth Environ. Sci.* **365**, 012004. <https://doi.org/10.1088/1755-1315/365/1/012004> (2019).
21. Hamada, H. M., Yahaya, F., Muthusamy, K. & Humada, A. Effect of incorporation POFA in cement mortar and desired benefits: a review. *IOP Conf. Ser. Earth Environ. Sci.* **365**, 012060. <https://doi.org/10.1088/1755-1315/365/1/012060> (2019).
22. Hamada, H. M. et al. Sustainable use of palm oil fuel ash as a supplementary cementitious material: A comprehensive review. *J. Build. Eng.* **40**, 102286. <https://doi.org/10.1016/j.jobe.2021.102286> (2021).
23. Altwair, N. M., Johari, M. A. M. & Hashim, S. F. S. Flexural performance of green engineered cementitious composites containing high volume of palm oil fuel ash. *Constr. Build. Mater.* **37**, 518–525. <https://doi.org/10.1016/j.conbuildmat.2012.08.003> (2012).
24. Khankhaje, E. et al. On blended cement and geopolymer concretes containing palm oil fuel ash. *Mater. Des.* **89**, 385–398. <https://doi.org/10.1016/j.matdes.2015.09.140> (2016).
25. Zeyad, A. M., Johari, M. A. M., Bunnori, N. M., Ariffin, K. S. & Altwair, N. M. Characteristics of Treated Palm Oil Fuel Ash and its Effects on Properties of High Strength Concrete. *Adv. Mat. Res.* **626**, 152–156. <https://doi.org/10.4028/WWW.SCIENTIFIC.NET/AMR.626.152> (2013).
26. Zeyad, A. M., Megat Johari, M. A., Tayeh, B. A. & Yusuf, M. O. Efficiency of treated and untreated palm oil fuel ash as a supplementary binder on engineering and fluid transport properties of high-strength concrete. *Constr. Build. Mater.* **125**, 1066–1079. <https://doi.org/10.1016/j.conbuildmat.2016.08.065> (2016).
27. Megat Johari, M. A., Zeyad, A. M., Muhammad Bunnori, N. & Ariffin, K. S. Engineering and transport properties of high-strength green concrete containing high volume of ultrafine palm oil fuel ash. *Constr. Build. Mater.* **30**, 281–288. <https://doi.org/10.1016/j.conbuildmat.2011.12.007> (2012).
28. Al-Mulali, M. Z., Awang, H., Abdul Khalil, H. P. S. & Aljoumaily, Z. S. The incorporation of oil palm ash in concrete as a means of recycling: A review. *Cem. Concr. Compos.* **55**, 129–138. <https://doi.org/10.1016/j.cemconcomp.2014.09.007> (2015).
29. Hasan, N. M. S. et al. Integration of rice husk ash as supplementary cementitious material in the production of sustainable high-strength concrete. *Materials* **15**(22), 8171. <https://doi.org/10.3390/ma15228171> (2022).
30. Ranjbar, N., Mehrali, M., Alengaram, U. J., Metselaar, H. S. C. & Jumaat, M. Z. Compressive strength and microstructural analysis of fly ash/palm oil fuel ash based geopolymer mortar under elevated temperatures. *Constr. Build. Mater.* **65**, 114–121. <https://doi.org/10.1016/j.conbuildmat.2014.04.064> (2014).
31. Hakeem, I. Y. et al. Properties of sustainable high-strength concrete containing large quantities of industrial wastes, nanosilica and recycled aggregates. *J. Mater. Res. & Technol.* **24**, 7444–7461. <https://doi.org/10.1016/j.jmrt.2023.05.050> (2023).
32. Altwair, N. M., Megat Johari, M. A., Saiyid Hashim, S. F. & Zeyad, A. M. Mechanical properties of engineered cementitious composite with palm oil fuel ash as a supplementary binder. *Adv. Mat. Res.* **626**, 121–125. <https://doi.org/10.4028/WWW.SCIENTIFIC.NET/AMR.626.121> (2013).
33. Mohammed, A. N., Johari, M. A. M., Zeyad, A. M., Tayeh, B. A. & Yusuf, M. O. Improving the Engineering and Fluid Transport Properties of Ultra-High Strength Concrete Utilizing Ultrafine Palm Oil Fuel Ash. *J. Adv. Concr. Technol.* **12**, 127–137. <https://doi.org/10.3151/JACT.12.127> (2014).
34. Zeyad, A. M., Megat Johari, M. A., Tayeh, B. A. & Yusuf, M. O. Pozzolanic reactivity of ultrafine palm oil fuel ash waste on strength and durability performances of high strength concrete. *J. Clean. Prod.* **144**, 511–522. <https://doi.org/10.1016/j.jclepro.2016.12.121> (2017).
35. Zeyad, A. M., Johari, M. A. M., Abutaleb, A. & Tayeh, B. A. The effect of steam curing regimes on the chloride resistance and pore size of high-strength green concrete. *Constr. Build. Mater.* **280**, 122409. <https://doi.org/10.1016/j.conbuildmat.2021.122409> (2021).
36. Zeyad, A. M. et al. Influence of steam curing regimes on the properties of ultrafine POFA-based high-strength green concrete. *J. Build. Eng.* **38**, 102204. <https://doi.org/10.1016/j.jobe.2021.102204> (2021).
37. Hamada, H. M., Al-Attar, A. A., Tayeh, B. & Yahaya, F. B. M. Optimizing the concrete strength of lightweight concrete containing nano palm oil fuel ash and palm oil clinker using response surface method. *Case Stud. Constr. Mater.* **16**, e01061. <https://doi.org/10.1016/j.cscm.2022.E01061> (2022).
38. Chindaprasirt, P., Homwuttitong, S. & Jaturapitakkul, C. Strength and water permeability of concrete containing palm oil fuel ash and rice husk-bark ash. *Constr. Build. Mater.* **21**, 1492–1499. <https://doi.org/10.1016/j.conbuildmat.2006.06.015> (2007).
39. Tangchirapat, W., Jaturapitakkul, C. & Chindaprasirt, P. Use of palm oil fuel ash as a supplementary cementitious material for producing high-strength concrete. *Constr. Build. Mater.* **23**, 2641–2646. <https://doi.org/10.1016/j.conbuildmat.2009.01.008> (2009).
40. Alsubari, B., Shafiqh, P. & Jumaat, M. Z. Utilization of high-volume treated palm oil fuel ash to produce sustainable self-compacting concrete. *J. Clean. Prod.* **137**, 982–996. <https://doi.org/10.1016/j.jclepro.2016.07.133> (2016).
41. Alsubari, B., Shafiqh, P. & Jumaat, M. Z. Development of self-consolidating high strength concrete incorporating treated palm oil fuel ash. *Materials* **8**(5), 2154–2173. <https://doi.org/10.3390/MA8052154> (2015).
42. Islam, M. M. U., Mo, K. H., Alengaram, U. J. & Jumaat, M. Z. Mechanical and fresh properties of sustainable oil palm shell lightweight concrete incorporating palm oil fuel ash. *J. Clean. Prod.* **115**, 307–314. <https://doi.org/10.1016/j.jclepro.2015.12.051> (2016).
43. Santhosh, K. G., Subhani, S. M. & Bahurudeen, A. Recycling of palm oil fuel ash and rice husk ash in the cleaner production of concrete. *J. Clean. Prod.* **354**, 131736. <https://doi.org/10.1016/j.jclepro.2022.131736> (2022).
44. Amin, M., Zeyad, A. M., Tayeh, B. A. & Saad Agwa, I. Effects of nano cotton stalk and palm leaf ashes on ultrahigh-performance concrete properties incorporating recycled concrete aggregates. *Constr. Build. Mater.* **302**, 124196. <https://doi.org/10.1016/j.conbuildmat.2021.124196> (2021).
45. Bashar, I. I., Alengaram, U. J. & Jumaat, M. Z. Enunciation of embryonic palm oil clinker based geopolymer concrete and its engineering properties. *Constr. Build. Mater.* **318**, 125975. <https://doi.org/10.1016/j.conbuildmat.2021.125975> (2022).
46. Adebayo Mujedu, K., Ab-Kadir, M. A. & Ismail, M. A review on self-compacting concrete incorporating palm oil fuel ash as a cement replacement. *Constr. Build. Mater.* **258**, 119541. <https://doi.org/10.1016/j.conbuildmat.2020.119541> (2020).

47. Liu, M. Y. J., Alengaram, U. J., Santhanam, M., Jumaat, M. Z. & Mo, K. H. Microstructural investigations of palm oil fuel ash and fly ash based binders in lightweight aggregate foamed geopolymer concrete. *Constr. Build. Mater.* **120**, 112–122. <https://doi.org/10.1016/j.conbuildmat.2016.05.076> (2016).
48. Zeyad, A. M., Tayeh, B. A., Saba, A. M. & Johari, M. A. M. Workability, setting time and strength of high-strength concrete containing high volume of palm Oil fuel ash, the open civil. *Eng. J.* **12**, 35–46. <https://doi.org/10.2174/1874149501812010035> (2018).
49. Alsubari, B., Shafigh, P., Ibrahim, Z., Alnahhal, M. F. & Jumaat, M. Z. Properties of eco-friendly self-compacting concrete containing modified treated palm oil fuel ash. *Constr. Build. Mater.* **158**, 742–754. <https://doi.org/10.1016/j.conbuildmat.2017.09.174> (2018).
50. Philip, J., Ismail, M. A. K. & Al-Subari, B. Strength and sulphate resistance of high performance concrete containing different fineness of Palm oil fuel ash. *IOP Conf. Ser. Mater. Sci. Eng.* **495**, 012100. <https://doi.org/10.1088/1757-899X/495/1/012100> (2019).
51. Feng, D. C. et al. Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. *Constr. Build. Mater.* **230**, 117000. <https://doi.org/10.1016/j.conbuildmat.2019.117000> (2020).
52. Ahmad, A. et al. Prediction of compressive strength of fly ash based concrete using individual and ensemble algorithm. *Materials* **14**(4), 794. <https://doi.org/10.3390/MA14040794> (2021).
53. Farooq, F., Ahmed, W., Akbar, A., Aslam, F. & Alyousef, R. Predictive modeling for sustainable high-performance concrete from industrial wastes: A comparison and optimization of models using ensemble learners. *J. Clean. Prod.* **292**, 126032. <https://doi.org/10.1016/j.jclepro.2021.126032> (2021).
54. Khan, M. A. et al. Compressive strength of fly-ash-based geopolymer concrete by gene expression programming and random forest. *Adv. Civ. Eng.* **2021**, 6618407. <https://doi.org/10.1155/2021/6618407> (2021).
55. Javed, M. F. et al. Applications of gene expression programming and regression techniques for estimating compressive strength of bagasse ash based concrete. *Crystals* **10**(9), 737. <https://doi.org/10.3390/CRYST10090737> (2020).
56. Getahun, M. A., Shitote, S. M. & Abiero Gariy, Z. C. Artificial neural network based modelling approach for strength prediction of concrete incorporating agricultural and construction wastes. *Constr. Build. Mater.* **190**, 517–525. <https://doi.org/10.1016/j.conbuildmat.2018.09.097> (2018).
57. Li, P., Zhao, H., Gu, J. & Duan, S. Dynamic constitutive identification of concrete based on improved dung beetle algorithm to optimize long short-term memory model. *Sci. Rep.* **14**, 1. <https://doi.org/10.1038/s41598-024-56960-z> (2024).
58. Huang, J. et al. Predicting the permeability of pervious concrete based on the beetle antennae search algorithm and random forest model. *Adv. Civ. Eng.* **2020**, 8863181. <https://doi.org/10.1155/2020/8863181> (2020).
59. Hao, Z. et al. Efficient hybrid multiobjective optimization of pressure swing adsorption. *Chem. Eng. J.* **423**, 130248. <https://doi.org/10.1016/j.cej.2021.130248> (2021).
60. Tanhadoust, A., Madhkhan, M. & Nehdi, M. L. Two-stage multi-objective optimization of reinforced concrete buildings based on non-dominated sorting genetic algorithm (NSGA-III). *J. Build. Eng.* **75**, 107022. <https://doi.org/10.1016/j.jobe.2023.107022> (2023).
61. Kellouche, Y., Tayeh, B. A., Chetbani, Y., Zeyad, A. M. & Mostafa, S. A. Comparative study of different machine learning approaches for predicting the compressive strength of palm fuel ash concrete. *J. Build. Eng.* **88**, 109187. <https://doi.org/10.1016/j.jobe.2024.109187> (2024).
62. Haque, M. A., Chen, B., Kashem, A., Qureshi, T. & Ahmed, A. A. M. Hybrid intelligence models for compressive strength prediction of MPC composites and parametric analysis with SHAP algorithm. *Mater. Today. Commun.* **35**, 105547. <https://doi.org/10.1016/j.mtcomm.2023.105547> (2023).
63. Abdulalim Alabdullah, A. et al. Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis. *Constr. Build. Mater.* **345**, 128296. <https://doi.org/10.1016/j.conbuildmat.2022.128296> (2022).
64. Nguyen, N. H., Abellán-García, J., Lee, S., Garcia-Castano, E. & Vo, T. P. Efficient estimating compressive strength of ultra-high performance concrete using XGBoost model. *J. Build. Eng.* **52**, 104302. <https://doi.org/10.1016/j.jobe.2022.104302> (2022).
65. Golafshani, E. M., Behnood, A. & Arashpour, M. Predicting the compressive strength of normal and high-performance concretes using ANN and ANFIS hybridized with Grey Wolf Optimizer. *Constr. Build. Mater.* **232**, 117266. <https://doi.org/10.1016/j.conbuildmat.2019.117266> (2020).
66. Kaloop, M. R., Kumar, D., Samui, P., Hu, J. W. & Kim, D. Compressive strength prediction of high-performance concrete using gradient tree boosting machine. *Constr. Build. Mater.* **264**, 120198. <https://doi.org/10.1016/j.conbuildmat.2020.120198> (2020).
67. Amjad, H., Khattak, M. M. H. & Khushnood, R. A. A simplified machine learning empirical model for biomimetic crack healing of bio-inspired concrete. *Mater. Today Commun.* **37**, 107063. <https://doi.org/10.1016/j.mtcomm.2023.107063> (2023).
68. Alahmari, T. S., Ullah, I. & Farooq, F. Machine learning models for estimating the compressive strength of rubberized concrete subjected to elevated temperature: Optimization and hyper-tuning. *Sustain. Chem. Pharm.* **42**, 101763. <https://doi.org/10.1016/j.scp.2024.101763> (2024).
69. Cakiroglu, C. et al. Explainable ensemble learning data-driven modeling of mechanical properties of fiber-reinforced rubberized recycled aggregate concrete. *J. Build. Eng.* **76**, 107279. <https://doi.org/10.1016/j.jobe.2023.107279> (2023).
70. De Domenico, D. et al. Compressive Strength Evaluation of Ultra-High-Strength Concrete by Machine Learning. *Materials* <https://doi.org/10.3390/MA15103523> (2022).
71. Duan, Z. H. & Poon, C. S. Properties of recycled aggregate concrete made with recycled aggregates with different amounts of old adhered mortars. *Mater. Des.* **58**, 19–29. <https://doi.org/10.1016/j.matdes.2014.01.044> (2014).
72. Zhao, Z. et al. Prediction of properties of recycled aggregate concrete using machine learning models: A critical review. *J. Build. Eng.* **90**, 109516. <https://doi.org/10.1016/j.jobe.2024.109516> (2024).
73. Kellouche, Y., Boukhatem, B., Ghrici, M. & Tagnit-Hamou, A. Exploring the major factors affecting fly-ash concrete carbonation using artificial neural network. *Neural. Comput. Appl.* **31**, 969–988. <https://doi.org/10.1007/S00521-017-3052-2/TABLES/8> (2019).
74. Shakr Piro, N., Mohammed, A., Hamad, S. M. & Kurda, R. Electrical resistivity-Compressive strength predictions for normal strength concrete with waste steel slag as a coarse aggregate replacement using various analytical models. *Constr. Build. Mater.* **327**, 127008. <https://doi.org/10.1016/j.conbuildmat.2022.127008> (2022).
75. Ahmed, H. U., Abdalla, A. A., Mohammed, A. S. & Mohammed, A. A. Mathematical modeling techniques to predict the compressive strength of high-strength concrete incorporated metakaolin with multiple mix proportions. *Clean. Mater.* **5**, 100132. <https://doi.org/10.1016/j.clema.2022.100132> (2022).
76. Jiang, H., Mohammed, A. S., Kazeroon, R. A. & Sarir, P. Use of the Gene-Expression Programming Equation and FEM for the High-Strength CFST Columns. *Appl. Sci.* **11**(21), 10468. <https://doi.org/10.3390/AP112110468> (2021).
77. D.K. Ismael Jaf, A. Abdalla, A.S. Mohammed, P.I. Abdulrahman, Rawaz Kurda, A.A. Mohammed, (2024) Hybrid nonlinear regression model versus MARS, MER, and ANN to evaluate the effect of the size and content of waste tire rubber on the compressive strength of concrete, *Heliyon* <https://doi.org/10.1016/j.heliyon.2024.e25997>
78. Zrar, Y. J., Abdulrahman, P. I., Sherwani, A. F. H., Younis, K. H. & Mohammed, A. S. Sustainable innovation in self-compacted concrete: Integrating by-products and waste rubber for green construction practices. *Structures* **62**, 106234. <https://doi.org/10.1016/j.istruc.2024.106234> (2024).

79. Emad, W. et al. Prediction of concrete materials compressive strength using surrogate models. *Structures* **46**, 1243–1267. <https://doi.org/10.1016/J.ISTRUC.2022.11.002> (2022).
80. Mohammed, A. et al. Prediction of compressive strength of concrete modified with fly ash: Applications of neuro-swarm and neuro-imperialism models. *Comput. Concr.* **27**(5), 489–512. <https://doi.org/10.12989/CAC.2021.27.5.489> (2021).
81. Das, P. & Kashem, A. Hybrid machine learning approach to prediction of the compressive and flexural strengths of UHPC and parametric analysis with shapley additive explanations. *Case Stud. Constr. Mater.* **20**, e02723. <https://doi.org/10.1016/J.CSCM.2023.E02723> (2024).
82. T. Guo, W. He, Z. Jiang, X. Chu, R. Malekian, Z. Li, An Improved LSSVM Model for Intelligent Prediction of the Daily Water Level, *Energies* 2019, Vol. 12, Page 112 12 (2018) 112. <https://doi.org/10.3390/EN12010112>.
83. Park, Y. S. & Lek, S. Artificial neural networks: Multilayer perceptron for ecological modeling. *Dev. Env. Model.* **28**, 123–140. <https://doi.org/10.1016/B978-0-444-63623-2.00007-4> (2016).
84. Sata, V., Jaturapitakkul, C. & Rattanashotinunt, C. Compressive strength and heat evolution of concretes containing palm oil fuel ash. *J. Mater. Civ. Eng.* **22**, 1033–1038. [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0000104](https://doi.org/10.1061/(ASCE)MT.1943-5533.0000104) (2010).
85. Tangchirapat, W. & Jaturapitakkul, C. Strength, drying shrinkage, and water permeability of concrete incorporating ground palm oil fuel ash. *Cem. Concr. Compos.* **32**, 767–774. <https://doi.org/10.1016/J.CEMCONCOMP.2010.08.008> (2010).
86. Hamada, H. M., Yahaya, F. M., Muthusamy, K., Jokhio, G. A. & Humada, A. M. Fresh and hardened properties of palm oil clinker lightweight aggregate concrete incorporating Nano-palm oil fuel ash. *Constr. Build. Mater.* **214**, 344–354. <https://doi.org/10.1016/J.CONBUILDMAT.2019.04.101> (2019).
87. Ul Islam, M. M., Mo, K. H., Alengaram, U. J. & Jumaat, M. Z. Durability properties of sustainable concrete containing high volume palm oil waste materials. *J. Clean. Prod.* **137**, 167–177. <https://doi.org/10.1016/J.JCLEPRO.2016.07.061> (2016).
88. Awal, A. S. M. A. & Shehu, I. A. Evaluation of heat of hydration of concrete containing high volume palm oil fuel ash. *Fuel* **105**, 728–731. <https://doi.org/10.1016/J.FUEL.2012.10.020> (2013).
89. Jaturapitakkul, C., Kiattikomol, K., Tangchirapat, W. & Saeting, T. Evaluation of the sulfate resistance of concrete containing palm oil fuel ash. *Constr. Build. Mater.* **21**, 1399–1405. <https://doi.org/10.1016/J.CONBUILDMAT.2006.07.005> (2007).
90. Sata, V., Jaturapitakkul, C. & Kiattikomol, K. Utilization of palm oil fuel ash in high-strength concrete. *J. Mater. Civ. Eng.* **16**, 623–628. [https://doi.org/10.1061/\(ASCE\)0899-1561\(2004\)16:6\(623\)](https://doi.org/10.1061/(ASCE)0899-1561(2004)16:6(623)) (2004).
91. (PDF) Influence of Palm Oil Fuel Ash on Properties of High-Strength Green Concrete, (n.d.). [https://www.researchgate.net/publication/333893210\\_Influence\\_of\\_Palm\\_Oil\\_Fuel\\_Ash\\_on\\_Properties\\_of\\_High-Strength\\_Green\\_Concrete](https://www.researchgate.net/publication/333893210_Influence_of_Palm_Oil_Fuel_Ash_on_Properties_of_High-Strength_Green_Concrete) (accessed October 8, 2024)
92. Wan Hassan, W. N. F. et al. Mixture optimization of high-strength blended concrete using central composite design. *Constr. Build. Mater.* **243**, 118251. <https://doi.org/10.1016/J.CONBUILDMAT.2020.118251> (2020).
93. Ranjbar, N., Behnia, A., Alsubari, B., Moradi Birgani, P. & Jumaat, M. Z. Durability and mechanical properties of self-compacting concrete incorporating palm oil fuel ash. *J. Clean. Prod.* **112**, 723–730. <https://doi.org/10.1016/J.JCLEPRO.2015.07.033> (2016).
94. Al-Mughanham, T., Aldhyani, T. H. H., Alsubari, B. & Al-Yaari, M. modeling of compressive strength of sustainable self-compacting concrete incorporating treated palm oil fuel ash using artificial neural network. *Sustainability* **12**(22), 9322. <https://doi.org/10.3390/SU12229322> (2020).
95. Sundus, K. I., Hammo, B. H., Al-Zoubi, M. B. & Al-Omari, A. Solving the multicollinearity problem to improve the stability of machine learning algorithms applied to a fully annotated breast cancer dataset. *Inform. Med. Unlocked* **33**, 101088. <https://doi.org/10.1016/J.IMU.2022.101088> (2022).
96. Garg, A. & Tai, K. Comparison of statistical and machine learning methods in modelling of data with multicollinearity. *Int. J. Model. Identif. Control.* **18**, 295–312. <https://doi.org/10.1504/IJMIC.2013.053535> (2013).
97. Drobnič, F., Kos, A. & Pustišek, M. On the Interpretability of machine learning models and experimental feature selection in case of multicollinear data. *Electronics* **9**(5), 761. <https://doi.org/10.3390/ELECTRONICS9050761> (2020).
98. Lindner, T., Puck, J. & Verbeke, A. Beyond addressing multicollinearity: Robust quantitative analysis and machine learning in international business research. *J. Int. Bus. Stud.* **53**, 1307–1314. <https://doi.org/10.1057/S41267-022-00549-Z/METRICS> (2022).
99. Javed, M. F., Siddiq, B., Onyelowe, K., Khan, W. A. & Khan, M. Metaheuristic optimization algorithms-based prediction modeling for titanium dioxide-Assisted photocatalytic degradation of air contaminants. *Results Eng.* **23**, 102637. <https://doi.org/10.1016/J.RINENG.2024.102637> (2024).
100. Alabduljabbar, H. et al. Predicting ultra-high-performance concrete compressive strength using gene expression programming method. *Case Stud. Constr. Mater.* **18**, e02074. <https://doi.org/10.1016/J.CSCM.2023.E02074> (2023).
101. Khan, M. & Javed, M. F. Towards sustainable construction: Machine learning based predictive models for strength and durability characteristics of blended cement concrete. *Mater. Today Commun.* **37**, 107428. <https://doi.org/10.1016/J.MTCOMM.2023.107428> (2023).
102. Khan, M. et al. Optimizing durability assessment: Machine learning models for depth of wear of environmentally-friendly concrete. *Results Eng.* **20**, 101625. <https://doi.org/10.1016/J.RINENG.2023.101625> (2023).
103. Alyami, M. et al. Predictive modeling for compressive strength of 3D printed fiber-reinforced concrete using machine learning algorithms. *Case Stud. Constr. Mater.* **20**, e02728. <https://doi.org/10.1016/J.CSCM.2023.E02728> (2024).
104. Alnahhal, A. M. et al. Synthesis of sustainable lightweight foamed concrete using palm oil fuel ash as a cement replacement material. *J. Build. Eng.* **35**, 102047. <https://doi.org/10.1016/J.JOBE.2020.102047> (2021).
105. Nazar, S. et al. Formulation of estimation models for the compressive strength of concrete mixed with nanosilica and carbon nanotubes. *Dev. the Built Environ.* **13**, 100113. <https://doi.org/10.1016/J.DIBE.2022.100113> (2023).
106. Jang, Y. et al. Multi-camera-based human activity recognition for human-robot collaboration in construction. *Sensors* **23**(15), 6997. <https://doi.org/10.3390/S23156997> (2023).
107. Amin, M. N. et al. Modeling Compressive Strength of Eco-Friendly Volcanic Ash Mortar Using Artificial Neural Networking. *Symmetry* **13**(11), 2009. <https://doi.org/10.3390/SYM13112009> (2021).
108. Alyami, M. et al. Estimating compressive strength of concrete containing rice husk ash using interpretable machine learning-based models. *Case Stud. Constr. Mater.* **20**, e02901. <https://doi.org/10.1016/J.CSCM.2024.E02901> (2024).
109. Aldrees, A., Khan, M., Taha, A. T. B. & Ali, M. Evaluation of water quality indexes with novel machine learning and SHapley Additive ExPlanation (SHAP) approaches. *J. Water Process. Eng.* **58**, 104789. <https://doi.org/10.1016/J.JWPE.2024.104789> (2024).
110. Alyami, M. et al. Application of metaheuristic optimization algorithms in predicting the compressive strength of 3D-printed fiber-reinforced concrete. *Dev. Built Environ.* **17**, 100307. <https://doi.org/10.1016/J.DIBE.2023.100307> (2024).
111. Aydogmus, H. Y. et al. A comparative assessment of bagging ensemble models for modeling concrete slump flow. *Comput. & concrete* **16**(5), 741–757. <https://doi.org/10.12989/CAC.2015.16.5.741> (2015).
112. Sharafati, A., Asadollah, S. B. H. S. & Al-Ansari, N. Application of bagging ensemble model for predicting compressive strength of hollow concrete masonry prism. *Ain Shams Eng. J.* **12**, 3521–3530. <https://doi.org/10.1016/J.ASEJ.2021.03.028> (2021).
113. Zhu, Y. et al. Predicting the splitting tensile strength of recycled aggregate concrete using individual and ensemble machine learning approaches. *Crystals (Basel)* **12**, 569. <https://doi.org/10.3390/CRYST12050569/S1> (2022).
114. Zhang, T. et al. Bagging-based machine learning algorithms for landslide susceptibility modeling. *Nat. Hazards* **110**(2), 823–846. <https://doi.org/10.1007/S11069-021-04986-1> (2021).
115. Bui, Q. A. T. et al. Prediction of interface shear stiffness modulus of asphalt pavement using bagging ensemble-based hybrid machine learning model. *Arab. J. Sci. Eng.* **48**, 13889–13900. <https://doi.org/10.1007/S13369-023-08014-1/METRICS> (2023).



116. Suykens, J. A. K. & Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **9**, 293–300. <https://doi.org/10.1023/A:1018628609742/METRICS> (1999).
117. Du, D., Jia, X. & Hao, C. A new least squares support vector machines ensemble model for aero engine performance parameter chaotic prediction. *Math. Probl. Eng.* **2016**, 4615903. <https://doi.org/10.1155/2016/4615903> (2016).
118. Kardani, M. N. & Baghban, A. Utilization of LSSVM strategy to predict water content of sweet natural gas. *Pet. Sci. Technol.* **35**, 761–767. <https://doi.org/10.1080/10916466.2016.1274758> (2017).
119. Azam, A. et al. Modeling resilient modulus of subgrade soils using LSSVM optimized with swarm intelligence algorithms. *Sci. Rep.* **12**(1), 1–20. <https://doi.org/10.1038/s41598-022-17429-z> (2022).
120. Kang, M. C., Yoo, D. Y. & Gupta, R. Machine learning-based prediction for compressive and flexural strengths of steel fiber-reinforced concrete. *Constr. Build. Mater.* **266**, 121117. <https://doi.org/10.1016/j.conbuildmat.2020.121117> (2021).
121. Abuodeh, O. R., Abdalla, J. A. & Hawileh, R. A. Assessment of compressive strength of ultra-high performance concrete using deep machine learning techniques. *Appl. Soft. Comput.* **95**, 106552. <https://doi.org/10.1016/j.asoc.2020.106552> (2020).
122. Chen, G., Suhail, S. A., Bahrami, A., Sufian, M. & Azab, M. Machine learning-based evaluation of parameters of high-strength concrete and raw material interaction at elevated temperatures. *Front. Mater.* **10**, 1187094. <https://doi.org/10.3389/FMATS.2023.1187094/BIBTEX> (2023).
123. Imran, M., Khushnood, R. A. & Fawad, M. A hybrid data-driven and metaheuristic optimization approach for the compressive strength prediction of high-performance concrete. *Case Stud. Constr. Mater.* **18**, e01890. <https://doi.org/10.1016/J.CSCM.2023.E01890> (2023).
124. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. & Data Min.* **10**(1145/2939672), 2939785 (2016).
125. Fan, J. et al. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Convers. Manag.* **164**, 102–111. <https://doi.org/10.1016/J.ENCONMAN.2018.02.087> (2018).
126. Kashem, A. & Das, P. Compressive strength prediction of high-strength concrete using hybrid machine learning approaches by incorporating SHAP analysis. *Asian J. Civ. Eng.* **24**, 3243–3263. <https://doi.org/10.1007/S42107-023-00707-0/METRICS> (2023).
127. Xu, Y. et al. Computation of high-performance concrete compressive strength using standalone and ensemble machine learning techniques. *Materials* **14**(22), 7034. <https://doi.org/10.3390/MA14227034> (2021).
128. Kellouche, Y., Boukhatem, B., Ghrici, M., Rebouh, R. & Zidol, A. Neural network model for predicting the carbonation depth of slag concrete. *Asian J. Civ. Eng.* **22**, 1401–1414. <https://doi.org/10.1007/S42107-021-00390-Z/METRICS> (2021).
129. Sahraoui, M. & Bouziani, T. ANN modelling approach for predicting SCC properties Research considering Algerian experience Part II Effects of aggregates types and contents. *J. Build. Mater. & Struct.* **8**(1), 63–71. <https://doi.org/10.34118/JBMS.V8I1.778> (2021).
130. Sada, S. O. & Ikpeseni, S. C. Evaluation of ANN and ANFIS modeling ability in the prediction of AISI 1050 steel machining performance. *Heliyon* **7**, e06136. <https://doi.org/10.1016/j.heliyon.2021.e06136> (2021).
131. H. Ayat, Y. Kellouche, M. Ghrici, B.B.-Adv.Comput. Des, undefined 2018, Compressive strength prediction of limestone filler concrete using artificial neural networks, Academia.EduH Ayat, Y Kellouche, M Ghrici, B BoukhatemAdv. Comput. Des, 2018•academia.Edu (n.d.). <https://www.academia.edu/download/92371460/acd0303004.pdf> (accessed October 8, 2024).
132. Fiorentini, N., Pellegrini, D. & Losa, M. Overfitting prevention in accident prediction models: Bayesian regularization of artificial. *Neural Netw.* **2677**, 1455–1470. <https://doi.org/10.1177/03611981221111367> (2022).
133. Piotrowski, A. P. & Napiorkowski, J. J. A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modelling. *J. Hydrol. (Amst.)* **476**, 97–111. <https://doi.org/10.1016/J.JHYDROL.2012.10.019> (2013).
134. Lohat, S., Jain, S. & Kumar, R. AROA: Adam Remora Optimization Algorithm and deep Q network for energy harvesting in Fog-IoV network. *Appl. Soft. Comput.* **136**, 110072. <https://doi.org/10.1016/J.ASOC.2023.110072> (2023).
135. Liu, B. & Liang, Y. Optimal function approximation with ReLU neural networks. *Neurocomputing* **435**, 216–227. <https://doi.org/10.1016/J.NEUCOM.2021.01.007> (2021).
136. Holland, J. H. Genetic Algorithms Computer programs that "evolve" in ways that resemble natural selection can solve complex problems even their creators do not fully understand. *Scientific American*, 66–72. <http://ia.deis.unibo.it/Courses/AI/fundamentalsAI2005-06/lucidi/seminari/roli/holland.GAIntro.pdf> (accessed October 8, 2024). (1992).
137. Saridemir, M. Genetic programming approach for prediction of compressive strength of concretes containing rice husk ash. *Constr. Build. Mater.* **24**, 1911–1919. <https://doi.org/10.1016/J.CONBUILDMAT.2010.04.011> (2010).
138. Pham, D. & Karaboga, D. Intelligent optimisation techniques: genetic algorithms, tabu search, simulated annealing and neural networks. Springer Science & Business Media. [https://books.google.com.pk/books?hl=en&lr=&id=FlndBwAAQBAJ&oi=fnd&pg=PA1&ots=KXDPBqOYQc&sig=joYuzFEFTFjuv\\_ksZICp3iF3vF4&redir\\_esc=y#v=onepage&q&f=false](https://books.google.com.pk/books?hl=en&lr=&id=FlndBwAAQBAJ&oi=fnd&pg=PA1&ots=KXDPBqOYQc&sig=joYuzFEFTFjuv_ksZICp3iF3vF4&redir_esc=y#v=onepage&q&f=false) (accessed October 8, 2024). (2012).
139. He, L. et al. A quasi-exponential distribution of interfacial voids and its effect on the interlayer strength of 3D printed concrete. *Addit. Manuf.* **89**, 104296. <https://doi.org/10.1016/J.ADDMA.2024.104296> (2024).
140. Karbasi, M. & Azamathulla, H. M. GEP to predict characteristics of a hydraulic jump over a rough bed. *KSCE J. Civ. Eng.* **20**, 3006–3011. <https://doi.org/10.1007/S12205-016-0821-X/METRICS> (2016).
141. Bagatur, T. & Onen, F. A predictive model on air entrainment by plunging water jets using GEP and ANN. *KSCE J. Civ. Eng.* **18**, 304–314. <https://doi.org/10.1007/S12205-013-0210-7/METRICS> (2014).
142. Althoey, F. et al. Machine learning based computational approach for crack width detection of self-healing concrete. *Case Stud. Constr. Mater.* **17**, e01610. <https://doi.org/10.1016/J.CSCM.2022.E01610> (2022).
143. Uddin, M. N., Ye, J., Deng, B. & L. zhi Li, K. Yu. Interpretable machine learning for predicting the strength of 3D printed fiber-reinforced concrete (3DP-FRC). *J. Build. Eng.* **72**, 106648. <https://doi.org/10.1016/J.JOBE.2023.106648> (2023).
144. Kumar, M. et al. Soft computing-based prediction models for compressive strength of concrete. *Case Stud. Constr. Mater.* **19**, e02321. <https://doi.org/10.1016/J.CSCM.2023.E02321> (2023).

## Acknowledgements

The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Research Project under grant number RGP2/176/45.

## Author contributions

1.Tariq Ali Project Administration Conceptualization Analysis 2.Kennedy C. Onyelowe Resources Supervision Funding 3.Muhammad Sarmad Mahmood Writing Original Paper 4.Muhammad Zeeshan Qureshi Review and Proofreading Methodology Editing 5.Nabil Ben Kahla Software Validation Supervision 6.Aïssa REZZOUG Data curation Proofreading 7.Ahmed deifalla Project Administration Supervision investigation.

## Declarations

## Competing interests

All authors confirm that they have no conflicts of interest.

## Ethics approval

All authors approve that the research was performed under all the ethical norms.

## Consent to publish

All authors consent to publishing this paper.

## Research limitation

This study is limited by its focus on a specific set of input parameters, without considering the influence of different concrete types, such as high-performance or self-compacting concrete. Additionally, external factors like temperature variations and the effects of POFA particle size were not explored, which may impact compressive strength predictions.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-89263-y>.

**Correspondence** and requests for materials should be addressed to K.C.O., M.Z.Q. or A.d.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025